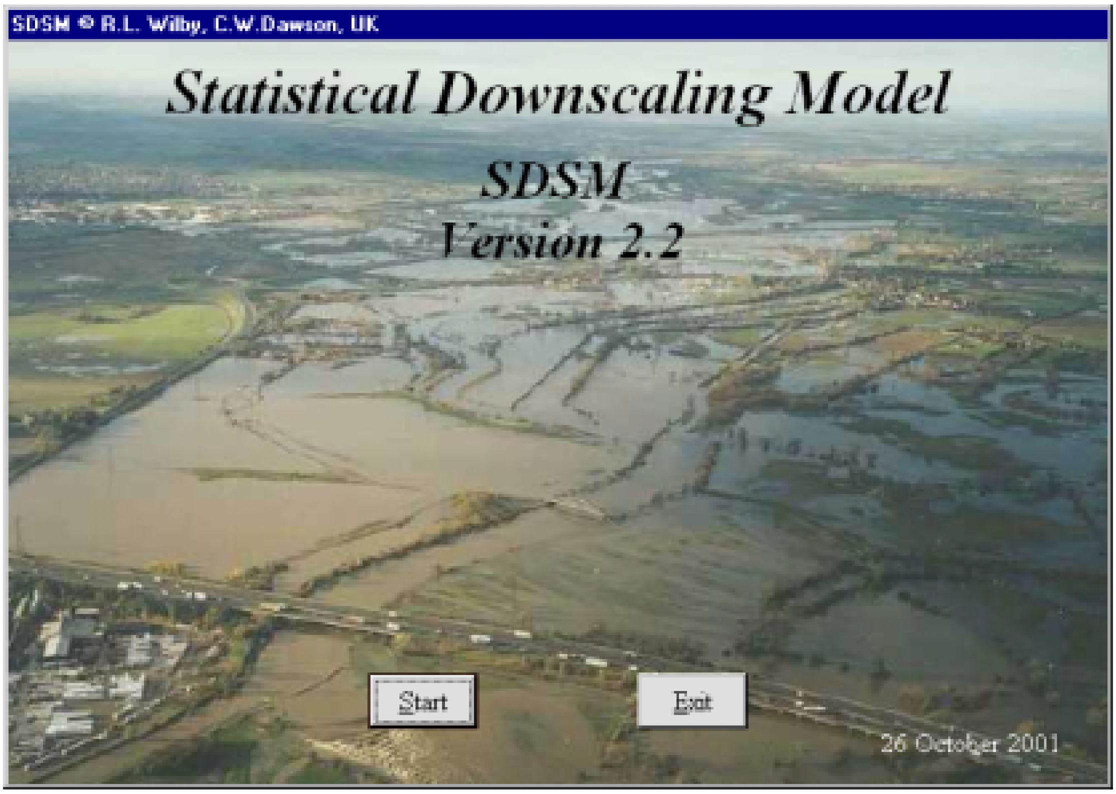
Using SDSM — A decision support tool for the assessment of regional climate change impacts



User Manual

Robert L. Wilby[1] and Christian W. Dawson[2]

26 October 2001

[1]Department of Geography, King's College London, Strand, WC2R 2LS, UK

[2]Department of Computer Science, Loughborough University, Leics., LE11 3TU, UK

Sponsors of SDSM

A Consortium for the Application of Climate Impact Assessments (ACACIA)

Canadian Climate Impacts Scenarios (CCIS) Project

National Centre for Risk Analysis and Options Appraisal, UK Environment Agency

Preface

General Circulation Models (GCMs) suggest that rising concentrations of greenhouse gases will have significant implications for climate at global and regional scales. Less certain is the extent to which meteorological processes at individual sites will be affected. So–called “downscaling” techniques are used to bridge the spatial and temporal resolution gaps between what climate modellers are currently able to provide and what impact assessors require.

This manual describes a decision support tool for assessing local climate change impacts using a robust statistical downscaling technique. SDSM (**S**tatistical **D**own**S**caling **M**odel) Version 2.2 facilitates the rapid development of multiple, low–cost, single–site scenarios of daily surface weather variables under current and future climate forcing. Additionally, the software performs ancillary tasks of data quality control and transformation, predictor variable pre–screening, automatic model calibration, basic diagnostic testing, statistical analyses and graphing of climate data.

In addition, the manual describes the UKSDSM archive, a set of daily predictor variables prepared for model calibration and downscaling at sites across the UK. This archive contains variables describing atmospheric circulation, thickness, stability and moisture content at several levels in the atmosphere, under climate conditions observed between 1961 and 2000. Equivalent predictor variables are provided for three GCM experiments of transient climate change between 1961 and 2099.

Application of SDSM is illustrated with respect to the downscaling of daily maximum temperature and precipitation scenarios for a hypothetical site at Blogsville, under current (1961–90) and future (2080–99) climate forcing.

Acknowledgements

Development of SDSM Version 2.2 was sponsored by the UK Environment Agency through the National Centre for Risk Analysis and Options Appraisal.

The authors gratefully acknowledge the support and technical advise of Nick Reynard (Environment Agency), Richenda Connell (UKCIP), Brian Waters (Environment Agency), and Iain Brown (UKCIP). The UKSDSM data archive was prepared by Ian Harris (Climate Research Unit). Owen Williams and Kit Macleod provided valuable feedback through beta–testing of the software.

Development of SDSM Version 2.1 was supported by the Canadian Climate Impacts Scenarios (CCIS) Group though the Climate Change Action Fund. Version 2.1 was also supported by A Consortium for the Application of Climate Impact Assessments (ACACIA) at the National Centre for Atmospheric Research (NCAR), Boulder, CO. NCAR is sponsored by the National Science Foundation.

NCEP Re–analysis data were provided by the NOAA-CIRES Climate Diagnostics Center, Boulder, Colorado, USA, from their Web site at http://www.cdc.noaa.gov/

The Climate Impacts LINK Project is funded by the UK Department of the Environment, Transport and the Regions (Contract Reference EPG 1/1/124).

CONTENTS

i Preface
ii Acknowledgements

0 TECHNICAL INFORMATION

- **0.1** System requirements
- **0.2** Installation

1 INTRODUCTION

- **1.1** Downscaling techniques
  - **1.1.1** Dynamical
  - **1.1.2** Weather typing
  - **1.1.3** Stochastic weather generators
  - **1.1.4** Regression
- **1.2** Relative skill of statistical and dynamical downscaling techniques
- **1.3** Manual outline

2 OVERVIEW OF SDSM STRUCTURE AND UKSDSM ARCHIVE

- **2.1** Key functions of SDSM
  - **2.1.1** Quality control and data transformation
  - **2.1.2** Selection of downscaling predictor variables
  - **2.1.3** Model calibration
  - **2.1.4** Weather generator
  - **2.1.5** Data analysis
  - **2.1.6** Graphing monthly statistics
  - **2.1.7** Scenario generation
- **2.2** UKSDSM data archive
- **2.3** UKSDSM predictors
- **2.4** SDSM file protocols
  - **2.4.1** UKSDSM file structure and nomenclature
  - **2.4.2** SDSM file name protocol

3 CAUTIONARY REMARKS

4 GETTING STARTED

- **4.1** Settings
- **4.2** Advanced settings

5 QUALITY CONTROL AND DATA TRANSFORMATION

- **5.1** Quality control
- **5.2** Data transformation

6 SELECTION OF DOWNSCALING PREDICTOR VARIABLES

6.1 Setup
6.2 Temporal variations in predictor strength
6.3 Correlation matrix
6.4 Scatterplot

7 MODEL CALIBRATION

7.1 File handling
7.2 Model type
7.3 Blogsville example

8 WEATHER GENERATOR

8.1 File handling
8.2 Ensemble size
8.3 Blogsville example

9 ANALYSIS OF OBSERVED AND DOWNSCALED DATA

9.1 Observed data
9.2 Downscaled data

10 GRAPHING MONTHLY STATISTICS

10.1 Line chart
10.2 Bar chart
10.3 Customizing charts

11 SCENARIO GENERATION

11.1 Check settings
11.2 Setup
11.2 Blogsville example (temperature)
11.3 Blogsville example (precipitation)

BIBLIOGRAPHY

APPENDICES

A.1 Enhancements since Version 2.1
A.2 Frequently asked questions

GLOSSARY

0. TECHNICAL INFORMATION

0.1 System requirements

Hardware requirements: PC Windows 98 or above.

Memory requirements: 5 MB RAM, 48 MB ROM.

Coding language: Visual Basic 6.0

Available since: 2001

0.2 Installation

Double click on the **SDSMInstall.exe** file.

A screen will appear explaining the installation process - click **OK** to continue.

The process will continue by creating and installing all necessary installation files to a directory on your C: drive called SDSM\InstallFiles (no other files will be deleted by this process).

Click on **Unzip** to continue.

Click **OK** when the files have extracted (unzipped).

The installation will continue automatically - click **OK** to start the installation process and follow the installation instructions on the screen.

You can now run SDSM from your Start Menu - select Programs, then SDSM, then SDSM (denoted by a cloud symbol).

1. INTRODUCTION

Even if global climate models in the future are run at high resolution there will remain the need to 'downscale' the results from such models to individual sites or localities for impact studies (DOE, 1996; p34).

General Circulation Models (GCMs) indicate that rising concentrations of greenhouse gases will have significant implications for climate at global and regional scales. Unfortunately, GCMs are restricted in their usefulness for local impact studies by their coarse spatial resolution (typically of the order 50,000 km^2) and inability to resolve important sub–grid scale features such as clouds and topography.

As a consequence, two sets of techniques have emerged as a means of deriving local–scale surface weather from regional–scale atmospheric predictor variables (Figure 1.1). Firstly, statistical downscaling is analogous to some of the techniques used for short–range numerical weather prediction since the 1970s. Secondly, Regional Climate Models (RCMs) simulate sub–GCM grid scale climate features dynamically using time–varying atmospheric conditions supplied by a GCM bounding a specified domain. Both approaches will continue to play a significant role in the assessment of potential climate change impacts arising from future increases in greenhouse–gas concentrations.

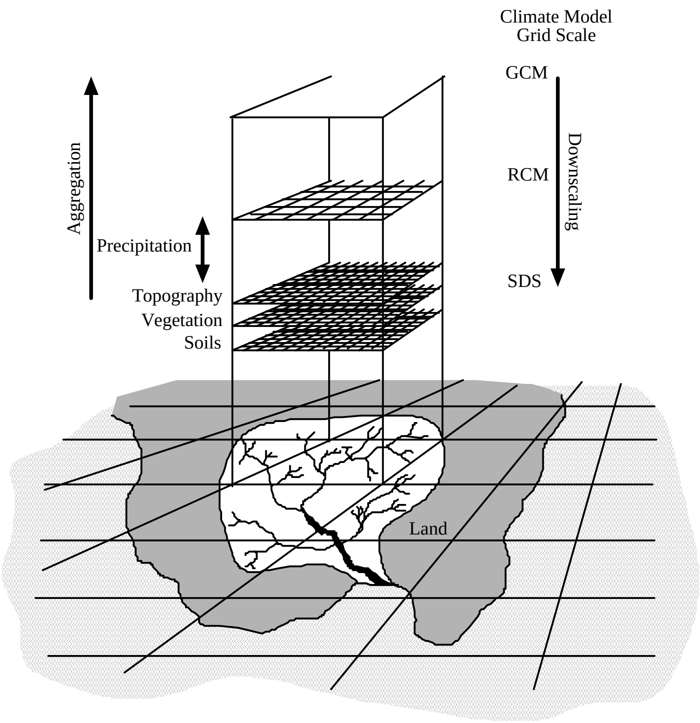


Figure 1.1 A schematic illustrating the general approach to downscaling.

Statistical downscaling methodologies have several practical advantages over dynamical downscaling approaches. In situations where low–cost, rapid assessments of localised climate change impacts are required, statistical downscaling currently represents the more promising option. In this manual we describe a software package,

and accompanying statistical downscaling methodology, that enables the construction of climate change scenarios for individual sites at *daily* time–scales, using grid resolution GCM output. The software is named SDSM (Statistical DownScaling Model) and is coded in Visual Basic 6.0.

As far as the authors are aware, SDSM is the first tool of its type offered to the broader climate change impacts community. Most statistical downscaling models are generally restricted in their use to specialist researchers and/or research establishments. Other software, although more accessible, produces relatively coarse regional scenarios of climate change (both spatially and temporally). For example, SCENGEN blends and re–scales user–defined combinations of GCM experiments, and then interpolates monthly climate change scenarios onto a 5º latitude × 5º longitude global grid. "Weather generators" — such as WGEN, LARS–WG or CLIGEN (see bibliography) — are widely used in the hydrological and agricultural research communities, but do not directly employ GCM output in the scenario construction processes.

Following a brief overview of downscaling techniques, we describe the structure and operation of SDSM with respect to seven tasks: 1) quality control and data transformation; 2) screening of potential downscaling predictor variables; 3) model calibration; 4) generation of ensembles of current weather data using *observed* predictor variables; 5) statistical analysis of observed data and climate change scenarios; 6) graphing model output; 7) generation of ensembles of *future* weather data using GCM–derived predictor variables. The key functions of SDSM will be illustrated using observed and climate model data for Blogsville, UK, comparing downscaled daily precipitation and temperature series for 1960–89 with 2080–99.

1.1 Downscaling techniques

The general theory, limitations and practice of downscaling have been discussed in detail elsewhere (see bibliography). Reviews typically group downscaling methodologies into four main types: a) dynamical climate modelling, b) synoptic weather typing, c) stochastic weather generation, or d) regression–based approaches. Each family of techniques is briefly described below.

1.1.1 Dynamical

Dynamical downscaling involves the nesting of a higher resolution Regional Climate Model (RCM) within a coarser resolution GCM. The RCM uses the GCM to define time–varying atmospheric boundary conditions around a finite domain, within which the physical dynamics of the atmosphere are modelled using horizontal grid spacings of 20–50 km. The main limitation of RCMs is that they are as computationally demanding as GCMs (placing constraints on the feasible domain size, number of experiments and duration of simulations). The scenarios produced by RCMs are also sensitive to the choice of boundary conditions (such as soil moisture) used to initiate experiments. The main advantage of RCMs is that they can resolve smaller–scale atmospheric features such as orographic precipitation or low–level jets better than the host GCM. Furthermore, RCMs can be used to explore the relative significance of different external forcings such as terrestrial–ecosystem or atmospheric chemistry changes.

1.1.2 Weather typing

Weather typing approaches involve grouping local, meteorological data in relation to prevailing patterns of atmospheric circulation. Future regional climate scenarios are constructed, either by re–sampling from the observed data distributions (conditional on the circulation patterns produced by a GCM), or by first generating synthetic sequences of weather patterns using Monte Carlo techniques and re–sampling from observed data. The main appeal of circulation–based downscaling is that it is founded on sensible linkages between climate on the large scale and weather at the local scale. The technique is also valid for a wide variety of environmental variables as well as multi–site applications. However, weather typing schemes are often parochial, a poor basis for downscaling rare events, and entirely dependent on stationary circulation–to–surface climate relationships. Potentially, the most serious limitation is that precipitation changes produced by changes in the frequency of weather patterns are seldom consistent with the changes produced by the host GCM (unless additional predictors such as atmospheric humidity are employed).

1.1.3 Stochastic weather generators

Stochastic downscaling approaches typically involve modifying the parameters of conventional weather generators such as WGEN or LARS–WG. The WGEN model simulates precipitation occurrence using two–state, first order Markov chains: precipitation amounts on wet days using a gamma distribution; temperature and radiation components using first–order trivariate autoregression that is conditional on precipitation occurrence. Climate change scenarios are generated stochastically using revised parameter sets scaled in direct proportion to the corresponding parameter changes in a GCM. The main advantage of the technique is that it can exactly reproduce many observed climate statistics and has been widely used, particularly for agricultural impact assessment. Furthermore, stochastic weather generators enable the efficient production of large ensembles of scenarios for risk analysis. The key disadvantages relate to the arbitrary manner in which precipitation parameters are adjusted for future climate conditions, and to the unanticipated effects that these changes may have on secondary variables such as temperature.

1.1.4 Regression

Regression–based downscaling methods rely on empirical relationships between local scale predictands and regional scale predictor(s). Individual downscaling schemes differ according to the choice of mathematical transfer function, predictor variables or statistical fitting procedure. To date, linear and non–linear regression, artificial neural networks, canonical correlation and principal components analyses have all been used to derive predictor–predictand relationships. The main strength of regression downscaling is the relative ease of application, coupled with their use of observable trans–scale relationships. The main weakness of regression–based methods is that the models often explain only a fraction of the observed climate variability (especially in precipitation series). In common with weather typing methods, regression methods also assume validity of the model parameters under future climate conditions, and regression–based downscaling is highly sensitive to the choice of predictor variables and statistical transfer function (see below).

Furthermore, downscaling future extreme events using regression methods is problematic since these phenomena, by definition, tend to lie at the limits or beyond the range of the calibration data set.

1.2 Relative skill of statistical and dynamical downscaling

Given the wide range of downscaling techniques (both dynamical and statistical) there is an urgent need for model comparisons using generic data sets and model diagnostics (see the Bibliography for exemplar studies). Until recently, this research was restricted to statistical–versus–statistical or dynamical–versus–dynamical model comparisons. However, a growing number of studies are undertaking statistical–versus–dynamical model comparisons and Table 1.1 summarises relative strengths and weaknesses that have emerged.

	Statistical downscaling	**Dynamical downscaling**
Strengths	• Station–scale climate information from GCM–scale output • Cheap, computationally undemanding and readily transferable • Ensembles of climate scenarios permit risk/ uncertainty analyses • Flexibility	• 10–50km resolution climate information from GCM–scale output • Respond in physically consistent ways to different external forcings • Resolve atmospheric processes such as orographic precipitation • Consistency with GCM
Weakness	• Dependent on the realism of GCM boundary forcing • Choice of domain size and location affects results • Requires high quality data for model calibration • Predictor–predictand relationships are often non–stationary • Choice of predictor variables affects results • Choice of empirical transfer scheme affects results • Low–frequency climate variability problematic	• Dependent on the realism of GCM boundary forcing • Choice of domain size and location affects results • Requires significant computing resources • Ensembles of climate scenarios seldom produced • Initial boundary conditions affect results • Choice of cloud/ convection scheme affects (precipitation) results • Not readily transferred to new regions

Table 1.1 Main strengths and weakness of statistical and dynamical downscaling.

The consensus of model inter–comparison studies is that dynamical and statistical methods display similar levels of skill at estimating surface weather variables under *current* climate conditions. However, because of recognised inter–variable biases in host GCMs, assessing the realism of *future* climate change scenarios produced by statistical downscaling methods is problematic. This is because uncertainties exist in *both* GCM and downscaled climate scenarios. For example, precipitation changes projected by the U.K. Meteorological Office coupled ocean–atmosphere model HadCM2, are over–sensitive to future changes in atmospheric humidity. Overall, the greatest obstacle to the successful implementation of both statistical and dynamical downscaling is the realism of the GCM output used to drive the schemes.

However, because of the parsimony and "low–tech" advantages of statistical downscaling methods over dynamical modelling (Table 1.1), a multiple regression–based method was chosen as the basis of the decision support tool, SDSM.

1.3 Manual outline

The rest of this manual is organised in seven main parts:

- **Section 2** provides a brief overview of the key operations in SDSM. For a complete description of the model specification, interested readers should refer to the articles listed in the Bibliography (see below). Descriptions of the UKSDSM data archive and associated file nomenclature are also provided in Section 2.

- **Section 3** provides a few cautionary remarks concerning the limitations of SDSM and appropriate usage. Users are strongly recommended to consider the issues raised here, before developing local scenarios using SDSM.

- **Sections 4** to **11** provide guidance on the practical implementation of the key functions in SDSM for downscaling regional climate change scenarios. Application of SDSM is illustrated using the hypothetical case of Blogsville.

- Next, a comprehensive **Bibliography** is supplied with follow–up references for more detailed discussions of the technical basis of SDSM, example applications and comparisons with other downscaling methods.

- Previous Users of SDSM Version 2.1 should refer to Appendix 1 for an update on the key **Enhancements** in Version 2.2.

- A trouble–shooting guide and outline of the most common pitfalls is provided in the form of a **Frequently Asked Questions** (FAQs) section in Appendix 2.

- Finally, definitions of commonly used technical terms related to statistical downscaling are provided in a **Glossary**.

2. OVERVIEW OF SDSM STRUCTURE AND UKSDSM ARCHIVE

Downscaling is justified whenever GCM (or RCM) simulations of variable(s) used for impacts modelling are unrealistic at the temporal and spatial scales of interest, either because the impact scales are below the climate model's resolution, or because of model deficiencies. At the same time the GCM must have demonstrable skill for large–scale variables that are strongly correlated with local processes. In practice, the choice of downscaling technique is also governed by the availability of archived observational and GCM data because both are needed to produce future climate scenarios.

The SDSM software reduces the task of statistically downscaling daily weather series into seven discrete processes (denoted by heavy boxes in Figure 2.1):

1) quality control and data transformation;
2) screening of predictor variables;
3) model calibration;
4) weather generation (observed predictors);
5) statistical analyses;
6) graphing model output;
7) scenario generation (climate model predictors).

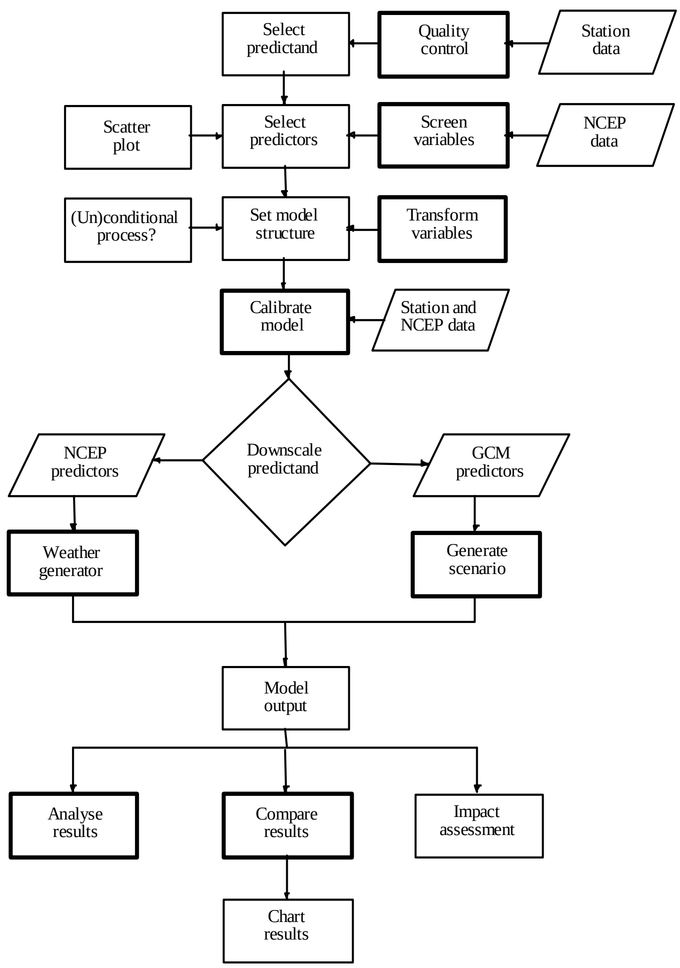


Figure 2.1 SDSM Version 2.2 climate scenario generation.

Full technical details of SDSM (and downscaling prototypes) are provided in the Bibliography. Within the taxonomy of downscaling techniques, SDSM is best described as a hybrid of the stochastic weather generator and regression–based methods. This is because large–scale circulation patterns and atmospheric moisture variables are used to linearly condition local–scale weather generator parameters (e.g., precipitation occurrence and intensity). Additionally, stochastic techniques are used to artificially inflate the variance of the downscaled daily time series to better accord with observations. To date, the downscaling algorithm of SDSM has been applied to a host of meteorological, hydrological and environmental assessments, as well as a range of geographical contexts including Europe, North America and Southeast Asia.

The following sections outline the software's seven core operations, along with the UKSDSM data archive and recommended file protocols.

2.1 Key functions of SDSM

As noted previously, SDSM performs seven key functions. The following paragraphs outline the purpose of each. Full technical explanation and User guidance are provided in Sections 4 to 11. **Next** and **Back** arrows (at the top of each screen) guide the User sequentially from one function to the next.

2.1.1 Quality control and data transformation

Few meteorological stations have 100% complete and/or fully accurate data sets. Handling of missing and imperfect data is necessary for most practical situations. Simple **Quality Control** checks in SDSM enable the identification of gross data errors, missing data codes and outliers prior to model calibration.

In many instances it may be appropriate to transform predictors and/or the predictand prior to model calibration. The **Transform** facility takes specified data files and applies selected transformations (e.g., logarithm, power, inverse, lag, binomial, etc).

2.1.2 Selection of downscaling predictor variables

Identifying empirical relationships between gridded predictors (such as mean sea level pressure) and single site predictands (such as station precipitation) is central to all statistical downscaling methods.

The main purpose of the **Screen Variables** operation is to assist the user in the selection of appropriate downscaling predictor variables. This remains one of the most challenging stages in the development of any statistical downscaling model since the choice of predictors largely determines the character of the downscaled climate scenario. The decision process is also complicated by the fact that the explanatory power of individual predictor variables varies both spatially and temporally. **Screen Variables** facilitates the examination of seasonal variations in predictor skill.

2.1.3 Model calibration

The **Calibrate Model** operation takes a user–specified predictand along with a set of predictor variables, and computes the parameters of multiple linear regression equations via the efficient dual simplex algorithm (forced entry method).

The User specifies the model structure: whether monthly, seasonal or annual sub–models are required; whether the process is unconditional or conditional. In unconditional models a direct link is assumed between the predictors and predictand (e.g., local wind speeds may be a function of regional airflow indices). In conditional models, there is an intermediate process between regional forcing and local weather (e.g., local precipitation amounts depend on the occurrence of wet–days, which in turn depend on regional–scale predictors such as humidity and atmospheric pressure).

2.1.4 Weather generator

The **Weather Generator** operation generates ensembles of synthetic daily weather series given observed (or NCEP re–analysis) atmospheric predictor variables. The procedure enables the verification of calibrated models (using independent data) and the synthesis of artificial time series representing current climate conditions.

The User selects a calibrated model and SDSM automatically links all necessary predictors to regression model weights. The User must also specify the period of record to be synthesised as well as the desired number of ensemble members. Synthetic time series are written to specific output files for later statistical analysis and/or impacts modelling.

2.1.5 Data analysis

Two operations provide basic descriptive statistics for observed data or SDSM derived scenarios. The **Analyse Other Data** screen takes as input observed climate data. The **Analyse Model Output** screen takes as input downscaled climate data produced by either the **Weather Generator** or **Generate Scenario** options.

In both cases, the User must specify the sub–period, and output file name. For model output, the ensemble member or mean, must also be specified. In return, SDSM displays a suite of diagnostics (chosen from the **Statistics** screen). The default statistics are monthly/ seasonal/ annual means, maxima, minima, sums and variances.

2.1.6 Graphing monthly statistics

The **Compare Results** screen enables the User to plot monthly statistics produced by the **Analyse Model Ouput/ Other Data** screen(s). Having specified the necessary input file, either bar or line charts may be chosen for display purposes. The graphing option allows simultaneous comparison of two data sets and hence rapid assessment of downscaled versus observed, or current versus future climate scenarios.

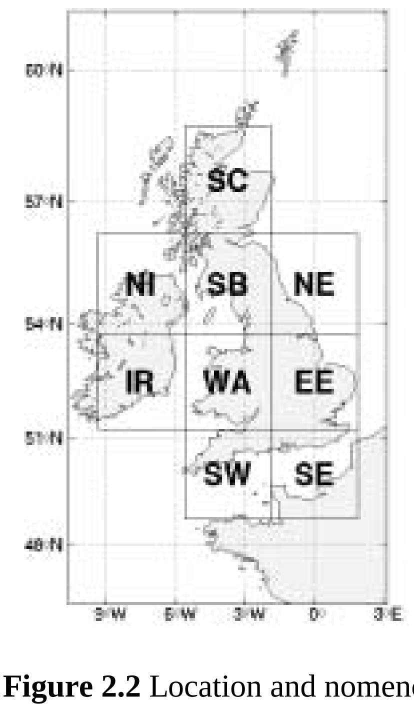
2.1.7 Scenario generation

Finally, the **Generate Scenario** operation produces ensembles of synthetic daily weather series given atmospheric predictor variables supplied by a climate model (either for current or future climate experiments), rather than observed predictors. The downscaling procedure is identical to that of the **Weather Generator** operation in all respects except that it may be necessary to specify a different convention for model dates and source directory for predictor variables. The input files for both the **Weather Generator** and **Generate Scenario** options need not be the same length as those used to obtain the regression weights during the calibration phase.

2.2 UKSDSM data archive

As Figure 2.1 indicates, the SDSM procedure commences with the preparation of coincident predictor and predictand data sets. Although the predictand is typically an individual daily weather series, obtained from meteorological observations at single stations (e.g., daily precipitation, maximum or minimum temperature, hours of sunshine, wind speed, etc.), the methodology is applicable to other environmental variables (e.g., air quality parameters, wave heights, snow cover, etc.). In any event, *these data must be supplied by the User in SDSM format* (see Section 2.4.2). This is single column, text only, beginning 1st January 1961, if necessary padded with the **Missing Data Identifier**.

Assembly of the candidate predictor suite is, by comparison, a far more involved process entailing data extraction, re–gridding and normalisation techniques. For this reason, SDSM is supplied with a prepared set of daily predictor variables for selected grid boxes covering the British Isles (Figure 2.2). The User simply locates the required grid box and data source from the UKSDSM archive. As Figure 2.2 shows the UK is represented by nine grid boxes each measuring 2.5º latitude by 3.75º longitude, corresponding to the grid co–ordinate system of the Hadley Centre’s coupled ocean–atmosphere GCMs (see below). Of the nine cells, six are land, and three are ocean. To obtain more realistic estimates of forcing over land areas that are represented by ocean grid boxes in the GCM, data from the two nearest land cells were averaged. For example, predictor variables for Southwest England (SW) are the average of data from the Wales (WA) and Southern England (SE) grid boxes.



SC	Scotland
NI[a]	Northern Ireland
SB	Scottish Boarders
NE[b]	Northeast England
IR	Ireland
WA	Wales
EE	Eastern England
SW[c]	Southwest England
SE	Southern England

[a]average of SB and IR
[b]average of SB and EE
[c]average of WA and SE

Figure 2.2 Location and nomenclature of the nine grid boxes used for downscaling.

For model calibration, the source is the National Centre for Environmental Prediction (NCEP) re–analysis data set. The data were re–gridded to conform to the system shown in Figure 2.2. All predictor variables (with the exception of the geostrophic wind direction, see below) were normalised with respect to the 1961 to 1990 average. However, daily predictors are also supplied for the period 1961–2000.

For downscaling future climate scenarios, the source is the UK Meteorological Office Hadley Centre's coupled ocean–atmosphere models HadCM2 and HadCM3 (see Bibliography). The UKSDSM data archive holds predictor variables from three experiments, each spanning 1961–2099. HadCM2GG is the ensemble mean for the HadCM2 greenhouse gas only experiment with CO_2 compounded annually by 1% per year. HadCM3SRESA2a and HadCM3SRESB2a are the two SRES scenarios A2a and B2a produced by HadCM3 with greenhouse gas, sulphate aerosol, and solar forcing.

2.3 UKSDSM predictors

Table 2.1 lists the daily predictor variables held in the UKSDSM data archive. Ideally, candidate predictor variables should be physically and conceptually sensible with respect to the predictand, strongly and consistently correlated with the predictand, and accurately modelled by GCMs. For precipitation downscaling, it is also recommended that the predictor suite contain variables describing atmospheric circulation, thickness, stability and moisture content. In practise, the choice of predictor variables is constrained by data availability from GCM archives. The predictors in Table 2.1, therefore, represent a compromise between maximum overlap between NCEP and HadCM2/HadCM3, as well as a range of choice for downscaling. It is envisaged that the list will be extended as further data are released from the Hadley Centre.

Daily variable	Code	NCEP 1961–2000	HadCM2 GG 1961–2099	HadCM3 SRES 1961–2099
Precipitation	prec		×	×
Maximum temperature	tmax		×	×
Minimum temperature	tmin		×	×
Mean temperature	temp	×		×
Mean sea level pressure	mslp	×		×
500 hPa geopotential height	p500	×		×
850 hPa geopotential height	p850	×		×
Near surface relative humidity	rhum	×		×
Relative humidity at 500 hPa height	r500	×		×
Relative humidity at 850 hPa height	r850	×		×
Near surface specific humidity	shum	×		×
Geostrophic airflow velocity	**_f	×	×	×
Vorticity	**_z	×	×	×
Zonal velocity component	**_u	×	×	×
Meridional velocity component	**_v	×	×	×
Wind direction	**th	×	×	×
Divergence	**zh	×	×	×

Table 2.1 Daily variables held in the UKSDSM data archive (denoted by ×). **Bold** type indicates variables that have **not** been normalised and are provided for comparative purposes. *Italics* indicate secondary (airflow) variables derived from pressure fields (surface, 500 and 850 hPa); refer to Jones *et al.* (1993).

2.4 SDSM file protocols

For convenience, the SDSM file protocol is described in two parts. Firstly, the file name system and file structure of the UKSDSM archive. Secondly, the meta–data and output files produced by SDSM more generally.

2.4.1 UKSDSM file structure and nomenclature

Figure 2.3 shows how the directory structure of the UKSDSM data archive relates to ancillary file systems in SDSM. The UKSDSM archive is organised into three levels. At the highest level are the data sources: presently NCEP, HadCM2 or HadCM3. At the second level, are the nine cells shown in Figure 2.2. At the third level, are files containing individual predictor variables.

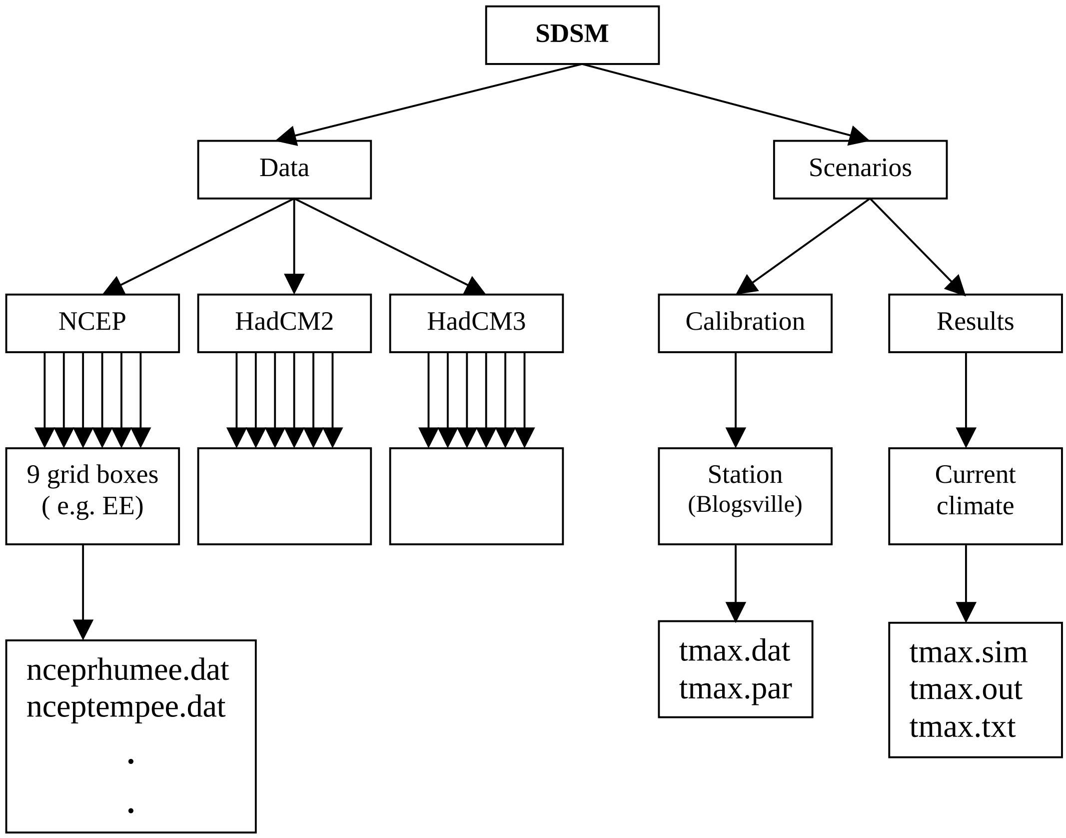


Figure 2.3 SDSM file structure with example file names (see Table 2.2 for definitions of file name extension).

Each file in the archive complies with a generic nomenclature of the form

[*source*] [*variable*] [*grid box*] . dat

The **source** is denoted by characters 1–4, the **variable** name by characters 5–8, and the **grid box** code by characters 9–10. All files have the extension *.dat. For example, the file name

nceprhumee.dat

indicates that the source is NCEP [*ncep*], the variable is near surface relative humidity [*rhum*], and the grid box is Eastern England [*ee*]. Similarly, the file name

h3b2p8_zsw.dat

indicates that the source is HadCM3, SRES scenario B2 [*h3b2*], the variable is vorticity computed at the 850 hPa geopotential height [*p8_z*], and the grid box is Southwest England [*sw*]. Alternatively, the file name

h2ggp_thsb.dat

indicates that the source is HadCM2, greenhouse gas only experiment [*h2gg*], the variable is surface wind direction [*p_th*], and the grid box is Scottish Boarders [*sb*].

2.4.2 SDSM file name protocol

With the above prerequisites in mind, Table 2.2 lists the file name extensions employed by SDSM, and Figure 2.3 shows the associated directory structures.

All input and output files are text only format. Individual predictor and predictand files (one variable to each file, time series data only) are denoted by the extension ***.dat**. Provided that the extension **.dat* is used, there are no restrictions on the nomenclature for station data provided by the User. The ***.PAR*** file records meta–data associated with the model calibration, model weights, and measures of "goodness–of–fit" (percentage explained variance and standard error of the model). The ***.SIM*** file records meta–data associated with every downscaled scenario (e.g., number of predictor variables, ensemble size, period, etc.), and the ***.OUT*** file contains an array of daily downscaled values (one column for each ensemble member, and one row for each day of the scenario). Finally, ***.TXT*** files are created whenever statistical analyses are undertaken in SDSM. These files record summary statistics for individual ensemble members or for the ensemble mean, and are accessed by bar/line chart options.

Extension	Explanation	Directory
*.DAT	Observed daily predictor and predictand files employed by the **Calibrate** and **Weather Generator** operations (input).	SDSM/Scenarios/ Calibration
*.PAR	Meta–data and model parameter file produced by the **Calibrate** operation (output) and used by the **Weather Generator** and **Generate Scenario** operations (input).	SDSM/Scenarios/ Calibration
*.SIM	Meta–data produced by the **Weather Generator** and **Generate Scenario** operations (output).	SDSM/Scenarios/ Results
*.OUT	Daily predictand variable file produced by the **Weather Generator** and **Generate Scenario** operations (output).	SDSM/Scenarios/ Results
*.TXT	Summary statistics produced by the **Analyse** operations (output).	SDSM/Scenarios/ Results

Table 2.2 SDSM file names and recommended directory structure.

3. CAUTIONARY REMARKS

SDSM is a Windows–based decision support tool for the rapid development of single–site, ensemble scenarios of daily weather variables under current and future regional climate forcing. Version 2.2 performs the tasks required to statistically downscale climate model output, namely: quality control of input data; screening of candidate predictor variables; model calibration; synthesis of current weather data; generation of future climate scenarios; basic statistical analyses; and graphing results. SDSM provides a robust and parsimonious technique of scenario construction that complements other methods (e.g., direct use of climate model output, dynamical downscaling, sensitivity analysis, etc.). Prospective Users should, therefore, consider the relative strengths and weaknesses of each category of downscaling to determine whether SDSM is most appropriate for the task in hand.

As far as the authors are aware, no comparable tool exists in the public domain. Nonetheless, the authors strongly caution that the software should **not** be used uncritically as a "black box". This is a very real danger when employing regression–based modelling techniques. Rather, the downscaling should be based upon physically sensible linkages between large–scale forcing and local meteorological response. Therefore, best practice demands rigorous evaluation of candidate predictor–predictand relationships using independent data. Furthermore, the local knowledge base is an invaluable source of information when determining sensible combinations of predictors.

Daily precipitation amounts at individual stations continue to be the most problematic variable to downscale, and research is ongoing to address this limitation. This arises because of the generally low predictability of daily precipitation amounts at local scales by regional forcing factors. The unexplained behaviour is currently modelled stochastically within SDSM by artificially inflating the variance of the downscaled series to accord better with daily observations. Even so, the model can produce unrealistic behaviour if the stochastic component is not properly handled. This, again, underlines the importance of independent testing of all model parameters against data withheld from model calibration.

Ultimately, however, the plausibility of all SDSM scenarios depends on the realism of the climate model forcing. Systematic biases in the mean and variance of GCM predictors can be reduced through normalisation with respect to a control period (as in the case of the UKSDSM predictors). Biases in large–scale patterns of atmospheric circulation in GCMs (e.g. shifts in the dominant storm track relative to observed data) or unrealistic inter–variable relationships are much harder to accommodate. Where possible, Users should not, therefore, restrict themselves to the use of a single GCM or emission scenario for downscaling. By applying multiple forcing scenarios (via different GCMs, ensemble members, time–slices, or emission pathways) better insight may be gained into the magnitude of these uncertainties.

Finally, the authors welcome constructive suggestions about the design or application of SDSM, particularly from the wider climate change impacts community. Please send your comments or questions to:

rob.wilby@kcl.ac.uk (downscaling methodology)
C.W.Dawson1@lboro.ac.uk (technical support)

4. GETTING STARTED

To start SDSM, click on the **Start** button on the Windows desktop, then on **Programs**, and then on **SDSM** (which will appear as a small rain cloud on the list of available programs). The following screen will appear:

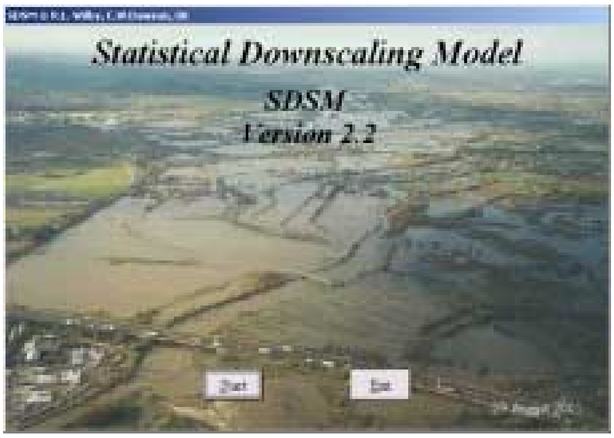


Figure 4.1 The SDSM "splash" screen.

Click on **Start** to continue to the SDSM main menu (Figure 4.2). If further information is required at any time, click on the **Help** button at the top of each screen (the User may then search the Help Contents by key word or task).

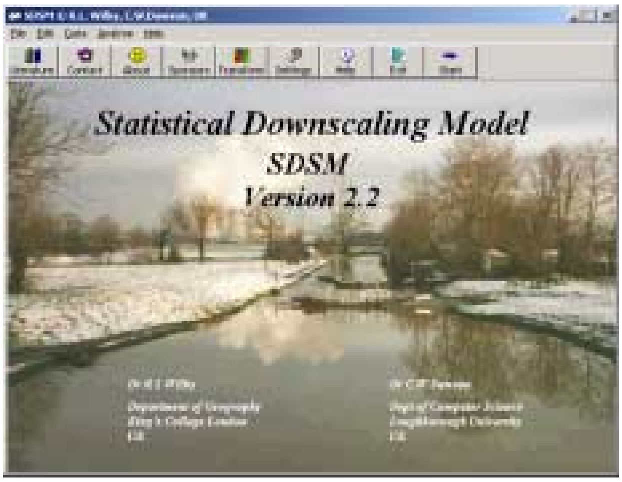


Figure 4.2 Main menu of SDSM.

Before downscaling, the User should check the date ranges, type and integrity of all input data. To establish the working environment click on the spanner symbol at the top of the main menu (or at the top of any other screen) to access the **Settings** screen (Figure 4.3).

4.1 Settings

The **Settings** screen may be accessed throughout SDSM. The following global preferences are available:

Year Length: The default allows 29 days in February every fourth year (i.e., leap years) and is generally used with observational data. The alternatives allow for different numbers of days in GCM data. For example, CGCM1 has 365 days and no leap years, whilst HadCM2 and HadCM3 have model years consisting of 360 days. Failure to set this parameter correctly can lead to system errors due to insufficient data.

Standard Start/End Date: Enter the global default start and end date for all input data. These dates will appear throughout the operation of SDSM, but may be updated from any screen.

Allow Negative Values: The default allows simulation of negative values by the downscaling model (e.g., for minimum temperature); deselection truncates values at zero (e.g., for sunshine hours).

Event Threshold: For some variables it is necessary to specify an event threshold. For example, when calibrating daily precipitation models, the parameter might be set to 0.3 mm/day to treat trace rain days as dry days. Similarly, the threshold for sunny versus cloudy days might be set at 1.0 hours/day to discriminate between overcast and sunny conditions.

Missing Data Identifier: This is the code assigned to missing data in all input series. Whenever SDSM encounters this code the value will be skipped (e.g., during model calibration, or calculation of summary statistics). The default is –999.

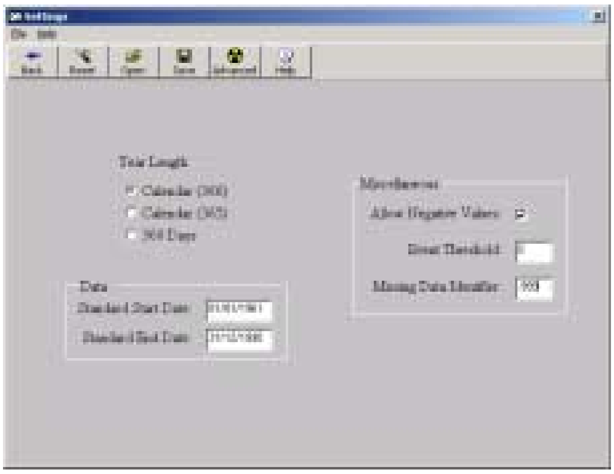


Figure 4.3 The **Settings** screen.

4.2 Advanced settings

The advanced settings are accessed from the **Settings** screen by clicking on the **Advanced** button at the top of the screen. The **Advanced Settings** screen allows the User to change and save further downscaling model preferences (Figure 4.4):

Model Transformation: Specifies the transformation applied to the predictand in conditional models. The default (None) is used whenever the predictand is normally distributed (this is often the case for daily temperature). The alternatives (Fourth root and Natural log) are used whenever data are skewed (as in the case of daily precipitation). The transformation type is recorded in *.PAR and *.SIM files to ensure that data are consistently handled during subsequent scenario and data analysis routines.

Variance Inflation: Controls the magnitude of variance inflation in downscaled daily weather variables. This parameter changes the variance by adding/reducing the amount of "white noise" applied to regression model estimates of the local process. The default value produces approximately normal variance inflation (prior to any transformation). Variance inflation is de–activated by setting the parameter to zero. Note that for Fourth root and Natural log **Model Transformations** (see above), this parameter also affects changes in the mean estimate.

Bias Correction: Compensates for any tendency to over– or under–estimate the mean of conditional processes by the downscaling model (e.g., mean daily rainfall totals). The default value is 1.0, indicating no bias correction.

Settings File: Locates standard and advanced settings held in a User defined reference file and directory. A new or updated settings file is created whenever the **Save** button is clicked at the top of the screen. The C:\SDSM.INI settings file is automatically loaded whenever SDSM is run.

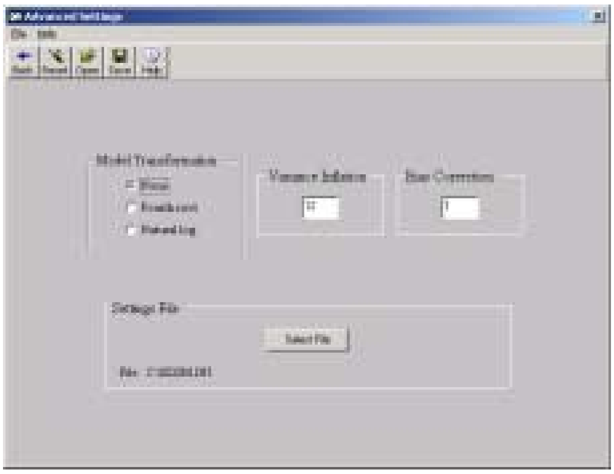


Figure 4.4 The **Advanced Settings** screen.

Press **Reset** at any time to reload the original settings, or **Back** to return to the **Settings** screen, followed by **Back** again to return to the last open screen.

5. QUALITY CONTROL AND DATA TRANSFORMATION

Few meteorological stations have complete and/or fully accurate data sets. Handling of missing and imperfect data is necessary for most practical situations. In some cases it is also necessary to transform data prior to model calibration. SDSM performs both quality control and data transformation.

5.1 Quality control

To check an input file for missing data and/or suspect values, click on the **Analyse** button at the top of the **Main Menu**, then select **Quality Control** from the drag down menu. The following screen will appear:

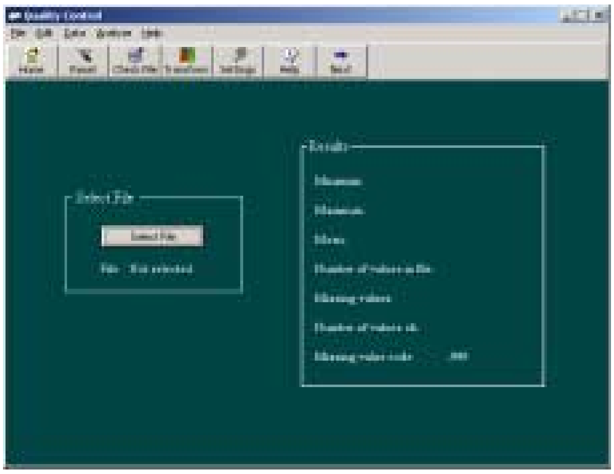


Figure 5.1 The **Quality Control** screen.

Click on the **Select File** button. An **Open** file window will appear – browse through until you have located the directory and file to be checked – in this example, the Blogsville maximum daily temperature, *TMAX.DAT*. Click on the required data file then on **Open**. To activate the quality control procedure, click on the **Check File** button at the top of the screen. The following confirmation will appear:

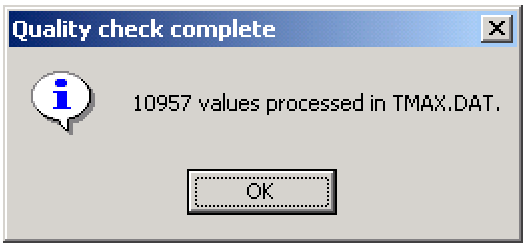


Figure 5.2 The **Quality check complete** dialogue box.

Click on the **OK** button to view the quality control information. In this example, there are 10957 values with no missing data (i.e., no missing value codes of –999 were detected). The data range from –6.7 to 34.8 ºC, with mean 13.1871ºC (see Figure 5.3). Click on the **Reset** button to clear the screen entries, or to perform a new quality check. Click on **Home** to return to the **Main Menu**, or on **Next** to proceed to **Screen Variables** (Section 6).

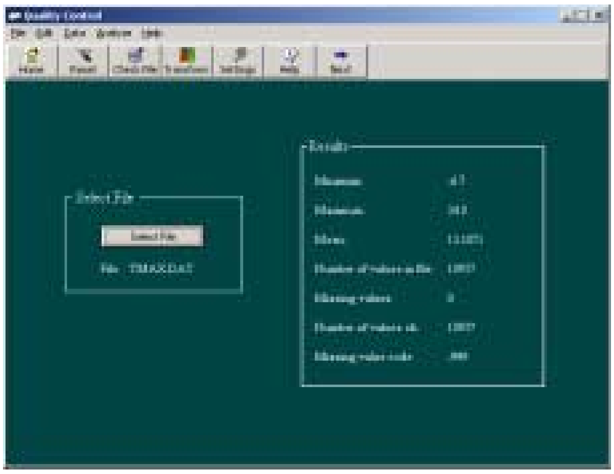


Figure 5.3 Results of the **Quality Control** check for *TMAX.DAT*.

5.2 Data transformation

To transform data, click on the **Transform** button at the top of the **Main Menu**. Alternatively click on the **Transform** button at the top of the **Quality Control** screen. In either case, the following screen will appear:

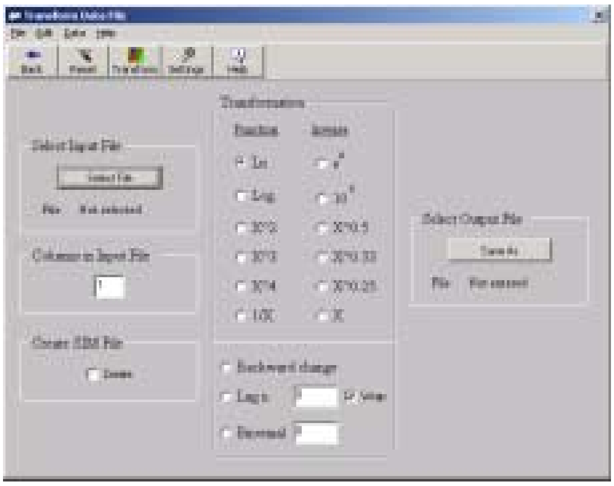


Figure 5.4 The **Transform Data File** screen.

Click on the **Select Input File** button. An **Open** file window will appear. Browse through until you have located the directory and file to be transformed – for example the surface vorticity over region XX, *ncepp__zxx.dat*. Click on the required file. If there is more than one column of input data (as in the case of an ensemble simulation produced by the **Weather Generator** or **Generate Scenario** functions) enter the appropriate number in the **Columns in Input File** box. To enable transformed data with multiple columns to be handled by the **Analyse Model Output** function (Section 9), check the box under **Create SIM File**.

Select the **Transformation** by checking the appropriate button. Available transformations include: natural logarithms and log10, squares, cubes, fourth powers, inversion, lag interval and binomial, together with the inverse transformations of the above where appropriate. If **Wrap** is selected (for **Lag n**) the last value is used as the

first value in the lag transformation; otherwise the **Missing Data Identifier** is inserted. The **Backward change** button is used to compute variable changes between successive days. All transformations can be applied to standard predictor variables prior to **Model Calibration** (Section 7), to produce non–linear regression models (e.g., use power transformations for polynomial models).

For region XX, select **Lag n**, enter "–1" in the box, and check the **Wrap** box (which will produce a lag–1 series of the variable with no missing data). Click on the **Select Output File** button. An **Open** file window will appear – browse through until the required directory is located, enter the **Filename** for transformed data, in this case *ncepzlagxx.dat* (i.e., vorticity on previous days), then click on **Save**. Note that, the name used for transformed files MUST comply fully with the protocol described in Section 2.4.1.

To activate the procedure, click on the **Transform** button at the top of the screen. The following confirmation will appear:

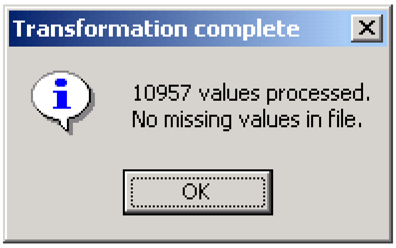


Figure 5.5 The **Transformation complete** dialogue box.

Click on the **OK** button to return to the **Transform Data File** screen. Click on the **Reset** button to clear the screen entries, or to perform a new transformation. Click on **Back** to return to the **Quality Control** screen, then on **Home** to return to the **Main Menu**.

6. SELECTION OF DOWNSCALING PREDICTOR VARIABLES

Identifying empirical relationships between gridded predictors (such as mean sea level pressure) and single site predictands (such as station precipitation) is central to all statistical downscaling methods. The purpose of the **Screen Variables** screen is to assist the User in the choice of appropriate downscaling predictor variables for model calibration (Section 7). SDSM performs three supporting tasks: seasonal correlation analysis, partial correlation analysis, and scatterplots. Ultimately, however, the User must decide whether or not the identified relationships are physically sensible for the site(s) in question.

To investigate potentially useful predictor–predictand relationships, click on the **Analyse** button at the top of any screen, then select **Screen Variables** from the drag down menu. The following screen will appear:

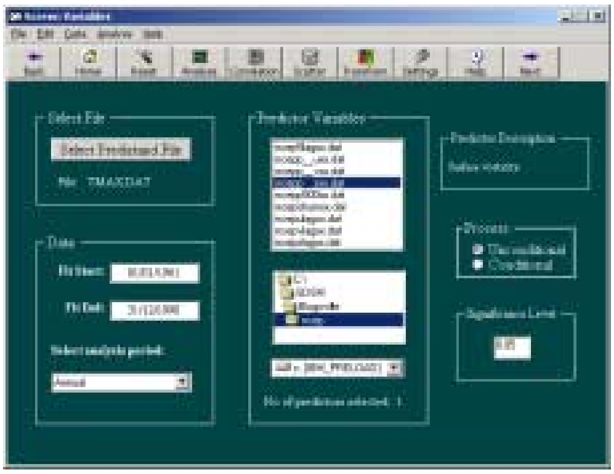


Figure 6.1 Illustration of **Screen Variables** using observed daily maximum temperatures (*TMAX*) for Blogsville, and NCEP predictors for region XX, 1961–1990.

6.1 Setup

The first step in the **Screen Variables** operation is the selection of the predictand and predictor files. The predictand file (e.g., observed daily maximum temperature, daily precipitation totals, etc.) must be supplied by the User, in SDSM format (see Section 2.4.2). Click on the **Select Predictand File** button. An **Open** file window will appear – browse through until the appropriate directory has been located. Click on the predictand data file – for example, in Figure 6.1 the maximum daily temperature at Blogsville, *TMAX.DAT*, located in C:\SDSM\Blogsville\observed.

Follow a similar procedure, locate and select the desired **Predictor Variables** by choosing the correct drive from the pull down window in the centre of the screen. The directories available on this drive will then appear in the window directly above the drive window. Browse through again until the appropriate directory is located. All *.dat files in this directory are then listed in the window above. To select a predictor, simply click on the file name – it will be highlighted in blue. A brief definition of the chosen variable is given in the **Predictor Description** window. To deselect a file, click on it again, and it will no longer be highlighted. The number of predictor variables chosen is shown beneath this window.

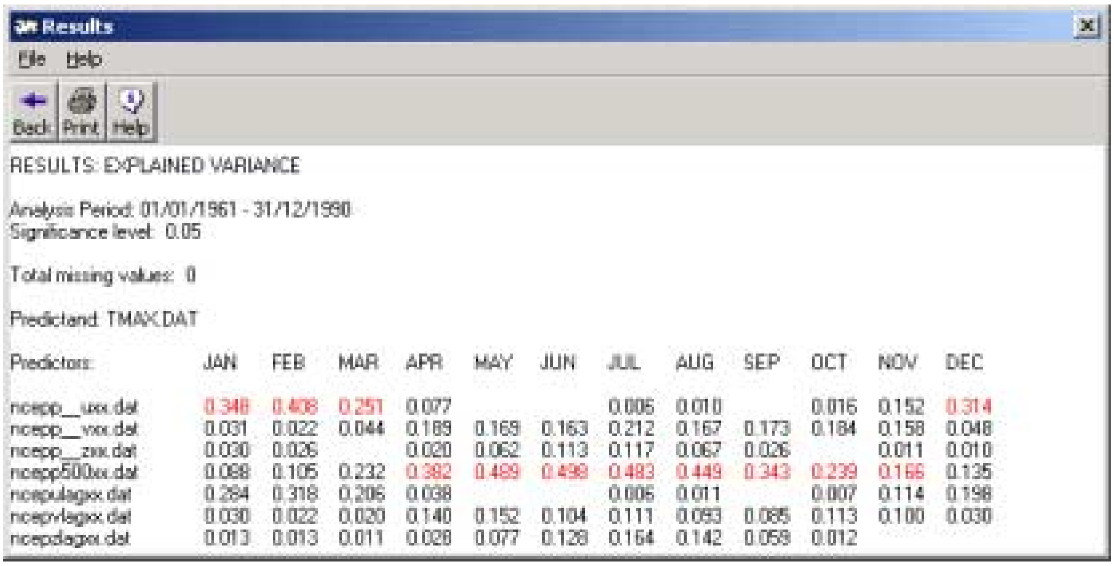
The **Data** menu on the left–hand side of the **Screen Variables** screen allows the start and end dates of the analysis period to be changed. The default dates are held in the **Settings** screen (see Section 4.1), in this case 1961–1990. If the start and end dates lie outside the permissible range, the User will be prompted to enter new values. The User must also choose the seasonal subset from the pull down window under **Select analysis period**. The available options are Annual (no seasonal sub–setting), Winter (December–February), Spring (March–May), Summer (June–August), Autumn (September–November), and individual months.

Two more actions are necessary before the analysis can take place. Firstly, the type of **Process** must be specified. If the predictor–predictand process is <u>not</u> regulated by an intermediate process (as in the case of maximum temperature) then click on **Unconditional**, otherwise select **Conditional** (as with precipitation where amounts depend on wet–day occurrence). Secondly, amend the **Significance Level** as required. This value is used to test the significance of predictor–predictand correlations. The default is $p<0.05$ (5%).

Once the above have been specified, SDSM is ready to analyse the chosen predictor–predictand relationship(s), for specified sub–period(s).

6.2 Temporal variations in predictor strength

The **Analyse** button is used to investigate the percentage of variance explained by specific predictand–predictor pairs. The strength of individual predictors often varies markedly on a month by month basis (e.g., Figure 6.2). The User should, therefore, be judicious concerning the most appropriate combination(s) of predictor(s) for a given season and predictand. The local knowledge base is also invaluable when determining sensible combinations of predictors.



RESULTS: EXPLAINED VARIANCE

Analysis Period: 01/01/1961 - 31/12/1990
Significance level: 0.05

Total missing values: 0

Predictand: TMAX.DAT

Predictors:	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
ncepp__uxx.dat	0.348	0.408	0.251	0.077			0.006	0.010		0.016	0.152	0.314
ncepp__vxx.dat	0.031	0.022	0.044	0.189	0.169	0.163	0.212	0.167	0.173	0.184	0.158	0.048
ncepp__zxx.dat	0.030	0.026		0.020	0.062	0.113	0.117	0.067	0.026		0.011	0.010
ncepp500xx.dat	0.098	0.105	0.232	0.392	0.489	0.498	0.483	0.449	0.343	0.239	0.166	0.135
ncepulagxx.dat	0.284	0.318	0.206	0.038			0.006	0.011		0.007	0.114	0.198
ncepvlagxx.dat	0.030	0.022	0.020	0.140	0.152	0.104	0.111	0.093	0.085	0.113	0.100	0.030
ncepzlagxx.dat	0.013	0.013	0.011	0.028	0.077	0.128	0.164	0.142	0.059	0.012		

Figure 6.2 The **Results** for the Blogsville example using NCEP predictors for region XX. The strongest correlation in each month is shown in red, indicating that the relationship between maximum temperature and *p500* and *p__u* are most important. Blanks represent insignificant relationships at the chosen **Significance Level**.

For the Blogsville example, select maximum daily temperatures as the predictand (*TMAX*), and the following NCEP predictor files: *p500, p__u, p__v, and*

p__z (see Table 2.1). In addition, use the **Transform** facility (Section 5.2) to create lagged values (one–day) for the three surface airflow indices. The predictand does not depend on an intermediate occurrence process, so **Unconditional** is checked under the **Process** option. Use the default dates for the **Data** option (i.e., 1961–1990), and choose **Annual** under **Select Analysis Period**. Use the default **Significance Level** (i.e., 0.05) then click on the **Analyse** button at the top of the **Screen Variables** menu. The results in Figure 6.2 suggest that *p500* is a potentially useful predictor for April through October maximum temperature, and *p__u* for December through March.

6.3 Correlation matrix

The **Correlation** button is used to investigate inter–variable correlations for specified sub–periods (annual, seasonal or monthly). SDSM also reports partial correlations between the selected predictors and predictand. These statistics help to identify the amount of explanatory power that is unique to each predictor.

For the Blogsville example, use the same predictand, predictors and set–up as in Section 6.2. Then click on the **Correlation** button at the top of the **Screen Variables** menu. The results are shown in Figure 6.3.

Note that the correlation matrix and partial correlation results are sensitive to the choice of the **Event Threshold** (in **Settings**) if a **Conditional** process is specified in the **Screen Variables** screen (see Figure 6.1).

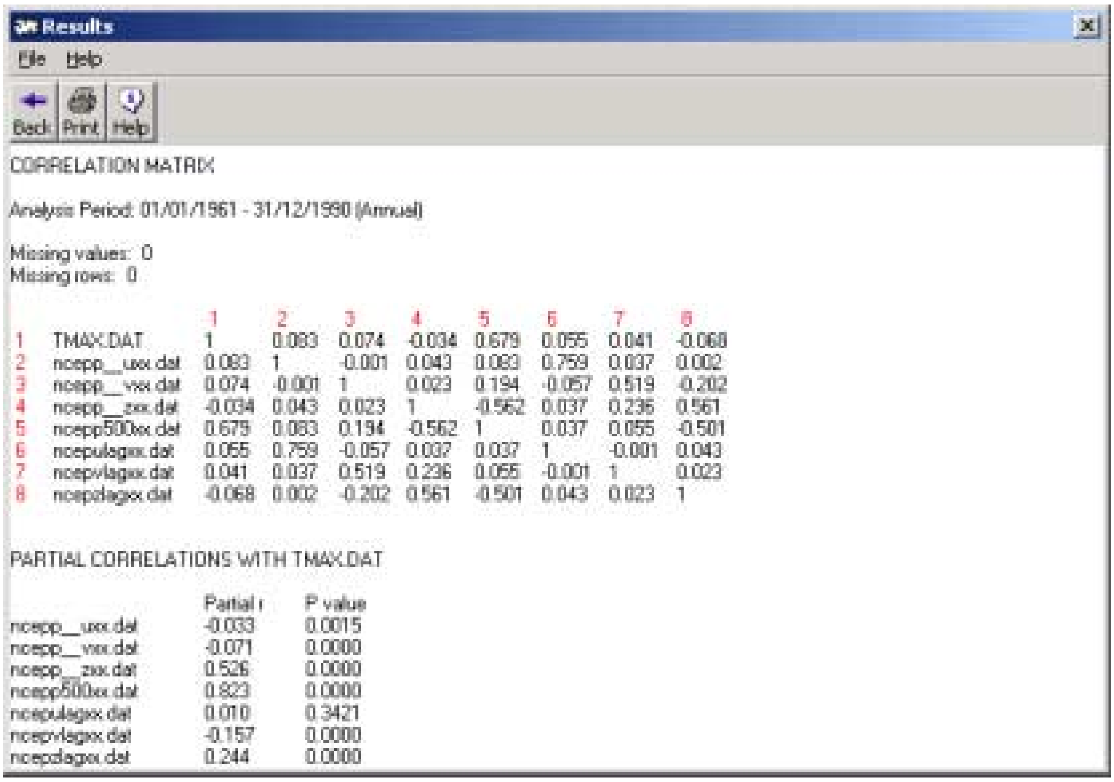


Figure 6.3 The **Results** for the Blogsville example. Partial correlations indicate that *p500* and *p__z* have the strongest association with *TMAX* once the influence of all other predictors has been removed.

6.4 Scatterplot

The **Scatter** button is used for visual inspections of inter–variable behaviour for specified sub–periods (annual, seasonal or monthly). The resultant scatterplot(s) indicate the nature of the association (linear, non–linear, etc.), whether or not data transformation(s) may be needed, and the importance of outliers.

For the Blogsville example, select *TMAX* as the predictand, *p__u* as the predictor file, and February under **Select analysis period** (following the results in Figure 6.2). Check that all other predictors have been deselected, and that **Unconditional** is selected under **Process**. (Note that if **Conditional** is selected, all predictand values less than or equal to the **Event Threshold** in **Settings** are excluded from the plot). Click on the **Scatter** button at the top of the **Screen Variables** menu. The results are shown in Figure 6.4.

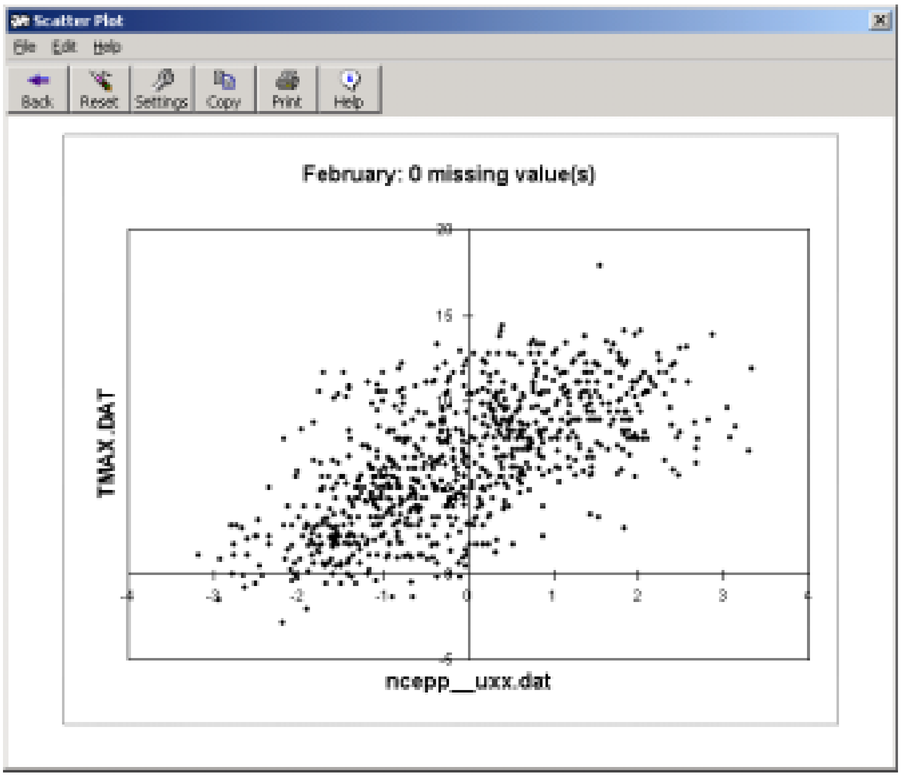


Figure 6.4 The **Scatterplot** for the Blogsville example, showing the association between *TMAX* and *p__u* in February. The results suggest that during February, higher maximum daily temperatures are associated with stronger westerly airflows.

The presentation quality of the **Scatterplot** may be customized, as required, by doubling clicking on any of the axis legends, titles or data points. Additional windows enable changes to be made to chart font, style, size, colour, etc.

To incorporate the **Scatterplot** in a Word document, first use the **Copy** button at the top of the screen, then in Word use **Paste Special (Picture)**.

7. MODEL CALIBRATION

The **Calibrate Model** process constructs downscaling models based on multiple linear regression equations, given daily weather data (the predictand) and regional–scale, atmospheric (predictor) variables. The parameters of the regression model are written to a standard format file with the extension *.PAR, along with meta–data recording details of the calibration period, model type, predictors used, etc.

The User specifies the model structure: whether monthly, seasonal or annual sub–models are required; whether the process is unconditional or conditional. In unconditional models a direct link is assumed between the predictors and predictand (e.g., local wind speeds may be a function of regional airflow indices). In conditional models, there is an intermediate process between regional forcing and local weather (e.g., local precipitation amounts depend on wet–/dry–day occurrence, which in turn depend on regional–scale predictors such as humidity and atmospheric pressure). Furthermore, it is possible to apply standard transformations to the predictand in conditional models (see Section 4.2), and/or to specific predictors (see Section 5.2).

To access the model building facility, click the **Analyse** button at the top of any screen, then select **Calibrate Model** from the drag down list. Alternatively, click on the **Next** button at the top of the **Screen Variables** screen. The following screen will appear:

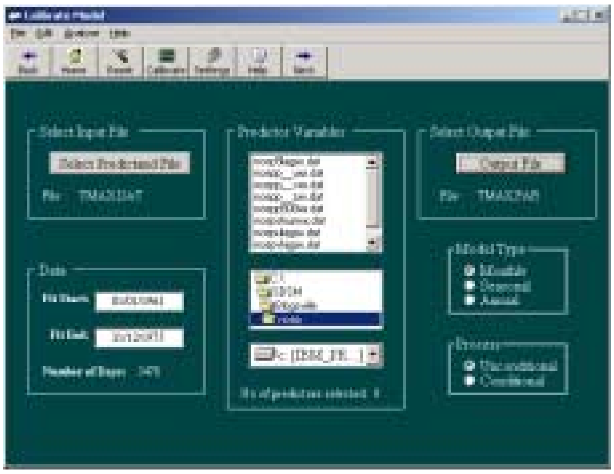


Figure 7.1 The **Calibrate Model** screen.

7.1 File handling

To begin model building, click on the **Select Predictand File** button in the top left–hand corner. An **Open** file window appears; browse through until the correct directory and file are reached, then click on the appropriate file name – for example, the maximum daily temperature at Blogsville, *TMAX.DAT*. The name of the file will then appear beneath the button.

Follow a similar procedure to locate and select the desired predictor variables by choosing the correct drive from the pull down window in the centre of the screen. The directories available on this drive will then appear in the window directly above the drive window. For example, the contents of C:\SDSM\Blogsville\ncep are shown in Figure 7.1. All *.DAT files in this directory are then listed in the window above. To select a predictor, simply click on the file name – it will be highlighted in blue. To

deselect a file, click on it again, and it will no longer be highlighted. The number of predictor variables chosen is shown beneath this window.

The **Data** menu on the left–hand side of the **Calibrate Model** screen allows the start and end dates of the analysis period to be changed. The default dates are held in the **Settings** screen (Section 4.1), in this case 1961–1990. If the start and end dates lie outside the permissible range, the User will be prompted to enter new values. Ideally, the model should be calibrated using a sub–set of available data, withholding the remainder for independent model validation (see Sections 8 and 9).

To specify the name of the output parameter (*.PAR) file, click on the **Output File** button in the upper right–hand corner. A **Save As** file window appears. For convenience, make sure that the parameter file is saved in the same directory as the predictand files, in this case, C:\SDSM\Blogsville\ncep. Type in an appropriate file name in the **File name** box then click on the **Save** button. The name of the parameter file will then be displayed beneath the **Output File** button, for example, TMAX.PAR.

7.2 Model type

To determine the temporal resolution of the downscaling model check either **Monthly**, **Seasonal** or **Annual** under the **Model Type** box. In **Monthly** models, different model parameters are obtained for each month. In **Seasonal** models, all months in the same season (e.g., December, January and February for winter) have the same model parameters. In **Annual** models, all months have the same parameters (i.e., there is no attempt to specify intra–annual variations in parameter values).

Next, indicate whether the downscaling process should be **Unconditional** or **Conditional** by checking the appropriate option in the **Process** box. Note that for conditional processes in which the distribution of predictand values is skewed, it is possible to apply a limited number of transformations in **Advanced Settings** (see Section 4.2). For example, the Fourth root might be selected for daily precipitation amounts. Finally, click the **Calibrate** button at the top of the screen.

7.3 Blogsville example

For the Blogsville example, five predictor files (*p500*, *p__u*, *vlag*, *p__z* and *zlag*) might be selected to downscale daily maximum temperatures, *TMAX* (see Figure 6.3). There is clearly a seasonal cycle in the regional forcing (Figure 6.2), so **Monthly** is checked in the **Model Type** box. The **Unconditional** option is checked in the **Process** box because a direct link is assumed to exist between the regional–scale predictors and local temperature. The date range in the **Data** menu is set at 1961 to 1975, ensuring that the second half of the data (i.e., 1976 to 1990) is retained for model validation.

Once the appropriate selections have been made, click on the **Calibrate** button. The process may take several seconds, and on completion a dialogue box will report the percentage of explained variance (R–squared value) and the Standard Error for the model. Figure 7.2 shows the results for the Blogsville example. In this case, about 55% of the variance in the local predictand is explained by regional forcing, and the standard error of the model is 2.2ºC.

Figure 7.2. The **Calibration Completed** dialogue box.

Click on the **OK** button to return to the **Calibrate Model** screen, then on **Home** to return to the **Main Menu**, on **Back** to return to **Screen Variables**, or on **Next** to proceed to the **Weather Generator**.

8. WEATHER GENERATOR

The **Weather Generator** operation produces ensembles of synthetic daily weather series given observed (or NCEP re–analysis) atmospheric predictor variables and regression model weights produced by the **Calibrate Model** operation (see Section 7). The **Weather Generator** enables the verification of calibrated models (assuming the availability of independent data) as well as the synthesis of artificial time series representative of current climate conditions.

To access this facility, click the **Analyse** button at the top of any screen, then select **Weather Generator** from the drag down list. Alternatively click on the **Next** button at the top of the **Calibrate Model** screen. The following screen will appear:

Figure 8.1 The **Weather Generator** screen.

8.1 File handling

The first step in the synthesis process is the selection of the appropriate parameter file. Click on the **Select Parameter File** button, in the top left–hand corner. An **Open** file window appears; browse through until the correct directory and file are reached, then click on the appropriate file name – for example, the parameters for maximum daily temperature at Blogsville, are held in *TMAX.PAR*. The name of the file will then appear beneath the button.

Next, specify the location of the predictor variable files by choosing the correct directory and drive from the window in the bottom left–hand corner of the screen under **Select Predictor Directory**.

To write synthetic data to a results file, it is necessary to select an appropriate directory and output file name. Click on the **Select Output File** button in the top right–hand corner. An **Open** file window appears; browse through until the correct directory is reached, then enter a suitable file name – for example, *TMAX.OUT*. The name of the file will then appear beneath the button.

Click on the **View Details** button, and the (predictand followed by predictor) files used in model calibration are listed in the window below. The **Record Start** date and available **Record Length** (number of days) are also displayed. The User must specify the (sub–) period required for weather generation using the **Synthesis Start** and **Sythesis Length** boxes, respectively.

The default values for **Synthesis Start** and **Sythesis Length** are used to simulate the period of record used for model calibration. If, however, model verification is to be undertaken using a set of independent data withheld from the calibration process, then the two values should be amended accordingly. If simulation of observed data based on the complete predictor record is needed, then the **Record Start** and **Record Length** values should be used.

8.2 Ensemble size

Decide how many ensembles of synthetic data are needed, up to a maximum of 100, and enter the appropriate value in the **Ensemble Size** box at the bottom right–hand corner of the screen (the default is 20). Individual ensemble members are considered equally plausible local climate scenarios realised by a common set of regional–scale predictors. The extent to which ensemble members differ depends on the relative significance of the deterministic and stochastic components of the regression models used for downscaling. For example, local temperatures are largely determined by regional forcing whereas precipitation series display more "noise" arising from local factors. As a general rule, the standard error of the ensemble mean, changes in proportion to the inverse square–root of the **Ensemble Size**.

Once the above selections have been completed, click the **Synthesize** button at the top of the menu. After a few seconds, the follow dialogue box will appear:

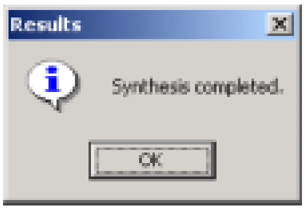


Figure 8.2 The synthesis completed dialogue box.

Click on **OK** to return to the **Weather Generator**, then on **Back** to return to the **Calibrate Model** screen, or on **Next** to proceed to **Analyse Model Output**.

8.3 Blogsville example

Selections for Blogsville are illustrated in Figure 8.1. In this example, the **Weather Generator** synthesized 20 runs of 15 years daily maximum temperature, using five regional–scale predictors. The data were synthesized using independent predictors withheld from model calibration (i.e., for the period 1976–1990).

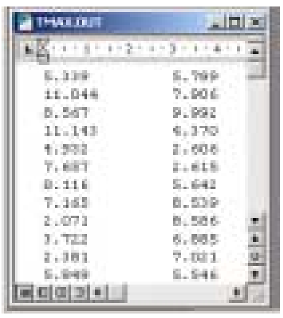
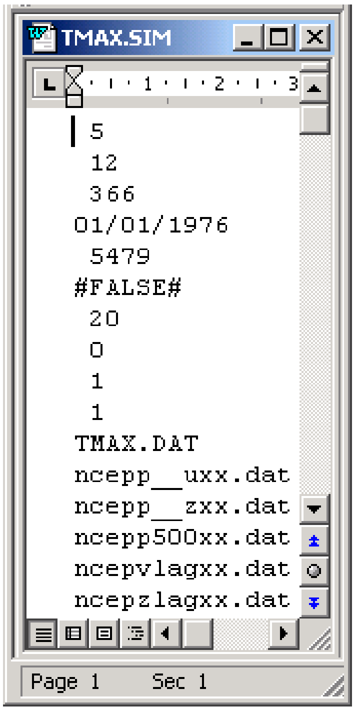


Figure 8.3 Example of the format of the simulated data file, *TMAX.OUT*.

Figure 8.3 shows the first twelve values of two ensemble members held in the *TMAX.OUT* file written to C:\SDSM\Blogsville\. Figure 8.4 shows the corresponding *TMAX.SIM* file, which contains meta–data associated with the synthesis. In both cases, the files have been opened in Word.


TMAX.SIM

5
12
366
01/01/1976
5479
#FALSE#
20
0
1
1
TMAX.DAT
ncepp__uxx.dat
ncepp__zxx.dat
ncepp500xx.dat
ncepvlagxx.dat
ncepzlagxx.dat

Page 1 Sec 1

```
[1] Number of predictor variables
[2] Number of regression models used
[3] Maximum number of days in a year
[4] Start date
[5] Number of days simulated
[6] Conditional or unconditional predictand
[7] Number of ensemble members
[8] Variance inflation parameter
[9] Transformation code
[10] Bias correction parameter
[11] Predictand file name
[12 onward] Predictor file name(s).
        "       "       "       "
        "       "       "       "
        "       "       "       "
        "       "       "       "
```

Figure 8.4 The *TMAX.SIM* file produced by the **Weather Generator** operation for the Blogsville example.

9. ANALYSIS OF OBSERVED AND DOWNSCALED DATA

Statistical tests of observed and downscaled weather data are handled in slightly different ways by SDSM. The **Analyse Other Data** operation produces summary statistics for single–site observed data series. The **Analyse Model Output** operation provides descriptive statistics for SDSM–derived scenarios, i.e., synthetic series produced by either the **Weather Generator** or **Generate Scenario** functions. However, common diagnostic tests are available for both observed and synthetic data. These statistics include the variable mean, maximum, minimum, variance, peaks above/below thresholds, percentiles, percent wet–days, and wet–/dry–day spell–lengths, computed on the basis of calendar month, seasonal and annual basis.

9.1 Observed data

To evaluate observed data, click on the **Analyse** button at the top of any screen, then select **Analyse Other Data** from the drag down list. The following screen will appear:

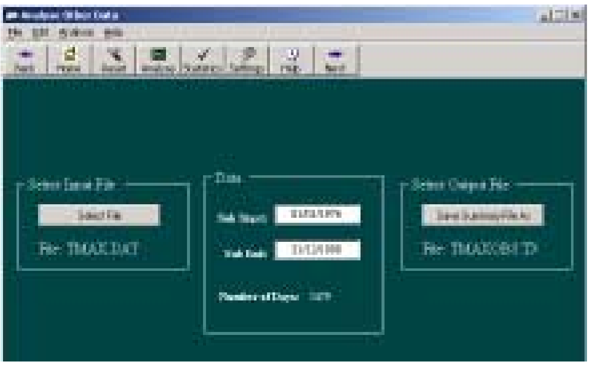


Figure 9.1 An example of the **Analyse Other Data** screen

The first step in the analysis process is the selection of the appropriate observed data (*.DAT) file. Click on the **Select File** button, on the left–hand side. An **Open** file window appears; browse through until the correct directory and file are reached, then click on the appropriate file name – for example, actual maximum daily temperatures at Blogsville, are held in *TMAX.DAT*. The name of the file will then appear beneath the button.

Next, specify the (sub–) period required for analysis using the **Sub Start** and **Sub End** windows, under the **Data** header. The default values are the **Standard Data Start Date** and **Standard Data End Date** held in the global **Settings** (Section 4).

To save the analysis results, it is necessary to select an appropriate directory and file name. Click on the **Save Summary File As** button on the right–hand side. An **Open** file window appears; browse through until the correct directory is reached, then enter a suitable file name – for example, *TMAXOBS.TXT*. The name of the file will then appear beneath the button.

The User is now ready to select required statistics. Click on the **Statistics** button at the top of the menu. The following screen will appear:

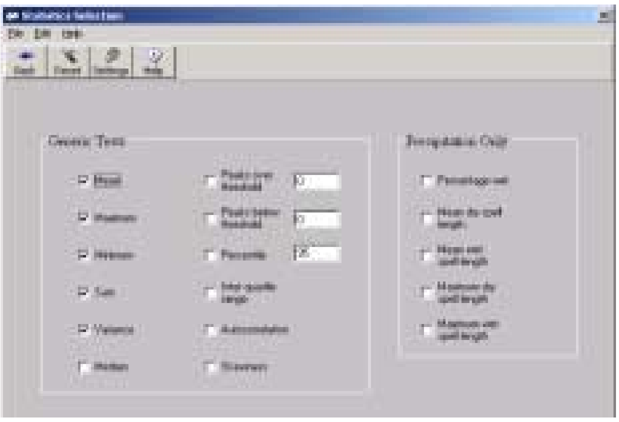
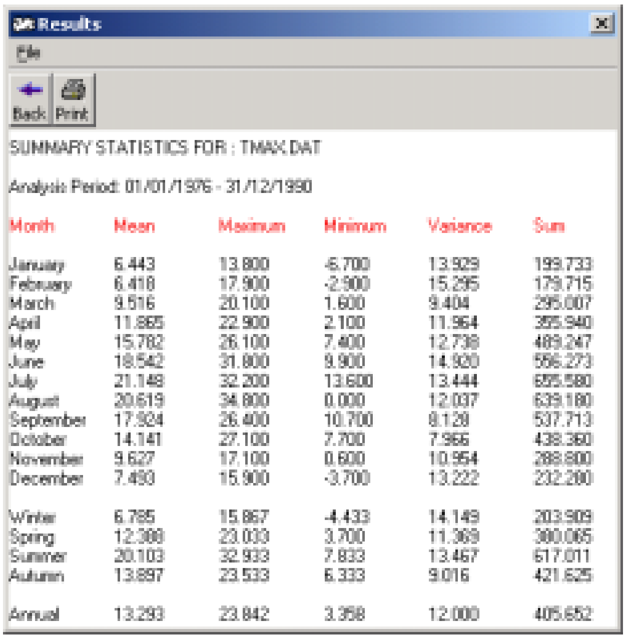


Figure 9.2 The **Statistics Selection** screen.

The screen is divided into two areas. The first, headed **Generic Tests**, lists statistics that may be applied to any variable (mean, maximum, minimum, sum, variance, median, peaks over/below threshold, percentile, inter–quartile range, autocorrelation and skewness). The second, headed **Precipitation Only**, lists statistics that are only applicable to daily precipitation series (percentage wet, mean dry–/wet–spell length, maximum dry–/wet–spell length). Note that the definition of a wet–day can be adjusted using the **Event Threshold** under **Settings**. The default is zero (i.e., wet–days are defined as all days with non–zero precipitation totals).

By checking the appropriate boxes, the User selects up to eight statistics for analysis. The defaults are the mean, maximum, minimum, sum, and variance. Click on **Back** to return to the **Analyse Other Data** screen.

Once all the above selections have been completed, click on the **Analyse** button at the top of the menu. After a few seconds, the **Results** screen will appear:



SUMMARY STATISTICS FOR : TMAX.DAT

Analysis Period: 01/01/1976 - 31/12/1990

Month	Mean	Maximum	Minimum	Variance	Sum
January	6.443	13.800	-6.700	13.929	199.733
February	6.418	17.900	-2.900	15.295	179.715
March	9.516	20.100	1.600	9.404	295.007
April	11.865	22.900	2.100	11.964	355.940
May	15.782	26.100	7.400	12.738	489.247
June	18.542	31.800	9.900	14.920	556.273
July	21.148	32.200	13.600	13.444	655.580
August	20.619	34.800	0.000	12.037	639.180
September	17.924	26.400	10.700	8.128	537.713
October	14.141	27.100	7.700	7.966	438.360
November	9.627	17.100	0.600	10.954	288.800
December	7.493	15.900	-3.700	13.222	232.280
Winter	6.785	15.867	-4.433	14.149	203.909
Spring	12.388	23.033	3.700	11.369	380.065
Summer	20.103	32.933	7.833	13.467	617.011
Autumn	13.897	23.533	6.333	9.016	421.625
Annual	13.293	23.842	3.358	12.000	405.652

Figure 9.3 Default statistics for observed daily maximum temperatures at Blogsville during the validation period 1976–1990.

The **Results** screen lists the name of the input file, along with the start and end dates of the analysis. Monthly, seasonal and annual mean statistics are listed beneath for the chosen tests. Comparison of the results obtained from the **Weather Generator** (see below) gives an indication of model skill. See Section 10 for visual comparisons of monthly statistics.

9.2 Downscaled data

To evaluate downscaled data, click on the **Analyse** button at the top of any screen, then select **Analyse Model Output** from the drag down list. Alternatively click on the **Next** button at the top of the **Weather Generator** screen. The following screen will appear:

Figure 9.4 An example of the **Analyse Model Output** screen.

The first step in the analysis process is the selection of the appropriate synthesis output (*.OUT) file. Click on the **Select File** button, in the top left–hand corner. An **Open** file window appears; browse through until the correct directory and file are reached, then click on the appropriate file name – for example, the results for maximum daily temperature at Blogsville, are held in *TMAX.OUT*. The name of the file will then appear beneath the button.

Next, click the **View Details** button, and details of the simulation experiment are displayed (the number of predictors, the number of seasonal models and year length, the scenario start date, number of days in the simulation, and the ensemble size). The User must specify the (sub–) period required for analysis using the **Analysis start date** and **Number of days** windows, beneath the **Data** header. The default values are for the full simulation period.

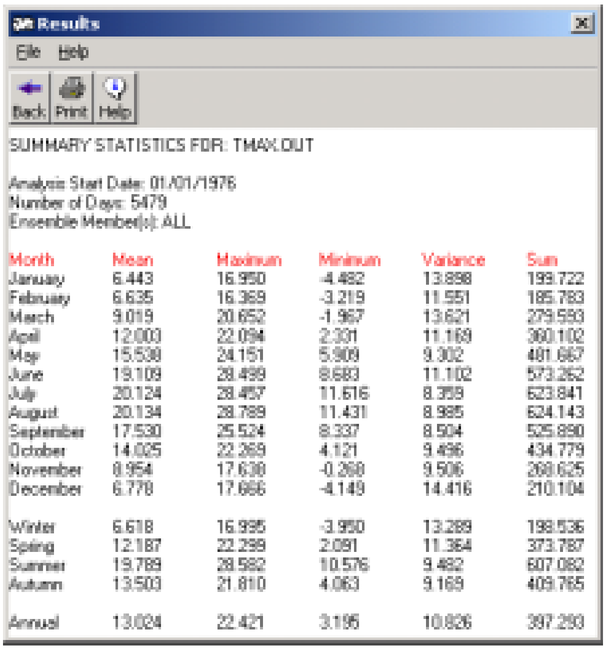
In the **Ensemble Size** section, check the **Use Ensemble Mean** box if average statistics are required from all ensemble members, or enter an individual ensemble member number (in this case, integers 1 to 20) in the **Ensemble Member** box. The default computes the ensemble mean of each statistic.

The User is now ready to select required statistics. Click on the **Statistics** button at the top of the menu and the screen shown in Figure 9.2 will appear. Check the appropriate tests, then click on **Back** to return to **Analyse Model Output** screen.

Finally, to save the results of the analysis, it is necessary to select an appropriate directory and file name. Click on the **Select Output File** button in the top right–hand corner. An **Open** file window appears; browse through until the correct

directory is reached, then enter a suitable file name – for example, *TMAX.TXT*. The name of the file will then appear beneath the button.

Once all the above selections have been completed, click on the **Analyse** button at the top of the menu. After a few seconds, the **Results** screen will appear:



SUMMARY STATISTICS FOR: TMAX.OUT

Analysis Start Date: 01/01/1976
Number of Days: 5479
Ensemble Member(s): ALL

Month	Mean	Maximum	Minimum	Variance	Sum
January	6.443	16.950	-4.482	13.898	199.722
February	6.635	16.369	-3.219	11.551	185.783
March	9.019	20.652	-1.967	13.621	279.593
April	12.003	22.094	2.331	11.169	360.102
May	15.538	24.151	5.909	9.302	481.667
June	19.109	28.499	8.683	11.102	573.262
July	20.124	28.457	11.616	8.359	623.841
August	20.134	28.789	11.431	8.985	624.143
September	17.530	25.524	8.337	8.504	525.890
October	14.025	22.269	4.121	9.496	434.779
November	8.954	17.638	-0.268	9.506	268.625
December	6.778	17.666	-4.149	14.416	210.104
Winter	6.618	16.995	-3.950	13.289	198.536
Spring	12.187	22.299	2.091	11.364	373.787
Summer	19.789	28.582	10.576	9.482	607.082
Autumn	13.503	21.810	4.063	9.169	409.765
Annual	13.024	22.421	3.195	10.826	397.293

Figure 9.5 Ensemble mean simulation results for the Blogsville example, using region XX predictors for the validation period 1976–1990.

This lists the name of the input file, the start date and number of days of the analysis, and the ensemble member (ALL indicates the ensemble mean, otherwise the integer entered in the **Ensemble Member** box appears). The monthly, seasonal and annual mean statistics are listed beneath. Comparison of the results with observed data (see Figure 9.3) gives an indication of model skill. See Section 10 for visual comparisons of monthly statistics.

10. GRAPHING MONTHLY STATISTICS

The **Compare Results** operation enables the User to plot monthly statistics produced by the **Analyse Model Ouput/ Other Data** operations (Section 9). Graphing options allow the comparison of two sets of results and hence rapid assessment of downscaled versus observed, or current versus future climate scenarios.

To access this facility, click the **Analyse** button at the top of any screen, then select **Compare Results** from the drag down list. The following screen will appear:

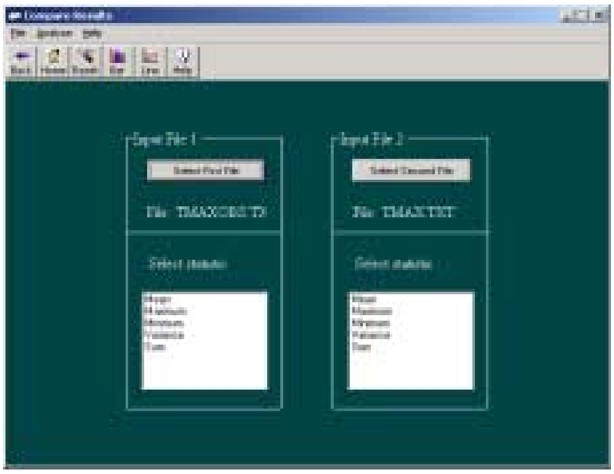


Figure 10.1 An illustration of the **Compare Results** screen.

10.1 Line chart

To choose a results (*.TXT) file, click on **Select First File** button. An **Open** file window appears; browse through until the correct directory and file are reached, then click on the appropriate file name – for example, observed statistics for maximum daily temperature at Blogsville, are held in *TMAXOBS.TXT* (see Section 9.1). The name of the file will then appear beneath the button, along with a list of available statistics. Repeat the process by clicking on the **Select Second File** button. Then click on the required statistic listed under **Select Statistic**. Finally, to show a line chart click on the **Line** button at the top of the screen:

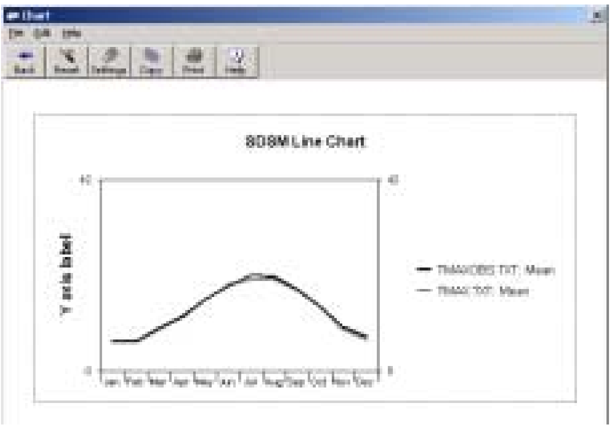


Figure 10.2 Example of the **Line** chart using observed (*TMAXOBS*) and downscaled (*TMAX*) monthly mean maximum daily temperatures at Blogsville, 1976–1990.

10.2 Bar chart

Alternatively, having selected the required files and statistics from each list (as in Section 10.1), click on the **Bar** button at the top of the **Compare Results** screen to produce a bar chart:

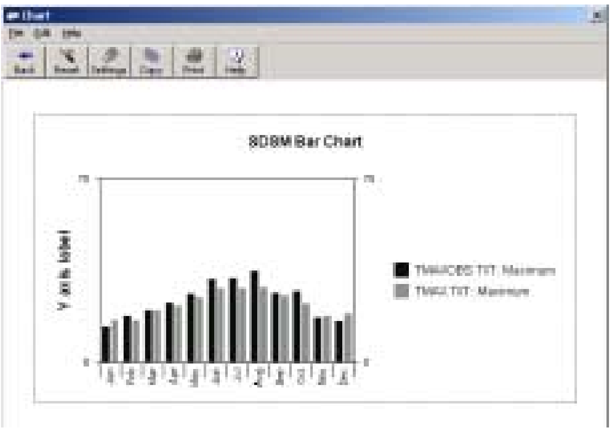


Figure 10.3 Example of the **Bar** chart using observed (*TMAXOBS*) and downscaled (*TMAX*) monthly absolute maximum daily temperatures at Blogsville, 1976–1990.

10.3 Customizing charts

To change (or remove) tick marks, y-axis labels, chart titles or y-axis maximum/minimum, in either the **Line** or **Bar** chart, click on the **Settings** button at the top of the screen. The following screen will appear:

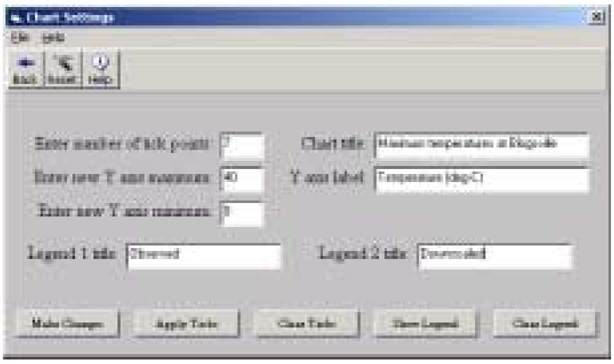


Figure 10.4 An illustration of the **Chart Settings** screen.

Enter the required values/text then click on the **Make Changes** button (to change text) and/or click on the **Apply Ticks** button to change tick marks. Similarly, click on the **Clear Ticks**, **Show Legend**, or **Clear Legend** buttons as required. Then click on the **Back** button to return to the plot.

To change the colour scheme of the lines or bars, double click on the object required. A **Color** palette will appear. Select the desired colour, then on **OK** to return to the chart. Similarly, to change the **Font**, double click on title and/or y-axis title. To change the position of text, single click then drag to the required location on the chart.

By applying the above design preferences, it is possible to change the bar chart in Figure 10.3 into the following:

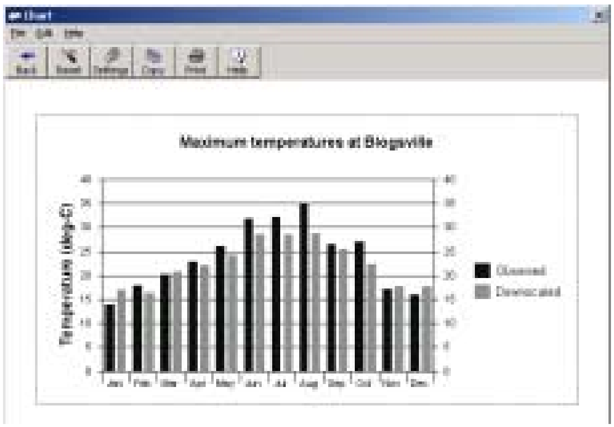


Figure 10.5 The same as Figure 10.3 but with customised bar colour scheme, tick marks, scale and legend.

Finally, to incorporate **Line** or **Bar** charts in a Word document, first use the **Copy** button at the top of the screen, then in Word use **Paste Special (Picture)**.

11. SCENARIO GENERATION

Generate Scenario produces ensembles of synthetic daily weather series given daily atmospheric predictor variables supplied by a GCM (either for current or future climate experiments). The GCM predictor variables must be normalised with respect to a reference period (or control run) and available for all variables used in model calibration. These tasks have already been undertaken (see Section 2.2).

The procedure is identical to that of the **Weather Generator** operation in all respects except that it may be necessary to specify a different convention for model dates and source directory for predictor variables. As in the case of the **Weather Generator** (see Section 8), input files for **Generate Scenario** need <u>not</u> be the same length as those used to obtain the regression weights during calibration.

To access this facility, click on the **Analyse** button at the top of any screen, then select **Generate Scenario** from the drag down list. The following screen appears:

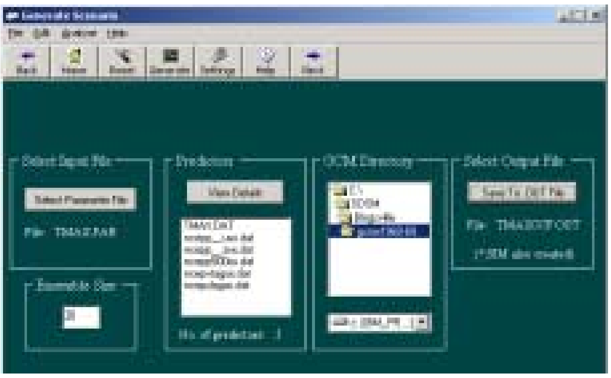


Figure 11.1 The **Generate Scenario** screen, showing selections for the Blogsville example using GCMX predictors representing the current (1960–1989) climate.

11.1 Check settings

Before starting scenario generation, it may be necessary to change some of the options in the **Settings** menu. Click on the **Settings** button at the top of the screen and check the appropriate **Year Length** box. Also, amend the **Standard Start/End Date** in line with the GCM data time–slices. For example, HadCM2, HadCM3 and GCMX have year lengths of 360 days, and for the Blogsville example, the period 1960–1989 was used to represent current climate forcing. Once necessary changes have been made to the **Settings**, click on **Back** to return to the **Generate Scenario** screen.

11.2 Setup

The first step in scenario generation is the selection of the appropriate downscaling model parameter file. Click on the **Select Parameter File** button, in the top left–hand corner. An **Open** file window appears; browse through until the correct directory and file are reached, then click on the appropriate file name – for example, the parameters for maximum daily temperature at Blogsville, are held in *TMAX.PAR*. The name of the file will then appear beneath the button.

Next, click on the **View Details** button, and the (predictand followed by predictor) files used in model calibration are listed in the window below. Then select the appropriate drive location and directory for the GCM predictors under the **GCM**

Directory header. For best practice, GCM predictors originating from different experiments or time–slices (e.g., 1960–1989 or 2080–2099) should be held in separate folders. This is because SDSM will load any files with the same predictor names (i.e., characters 5 to 8) as those used in model calibration (see Table 2.1).

As in the **Weather Generator** (Section 8), decide how many ensembles members are needed, up to a maximum of 100, and enter the appropriate integer in the **Ensemble Size** box on the left–hand side of the screen (the default is 20).

Finally, to save the scenario data to a results file, it is necessary to select an appropriate directory and file name. Click on the **Select Output File** button in the top right–hand corner. An **Open** file window appears; browse through until the correct directory is reached, then enter a suitable file name – for example, *TMAXCCF.OUT* (maximum temperature, current climate forcing). The name of the file will then appear beneath the button.

Once all the above selections have been completed, click on the **Generate** button at the top of the screen. After a few seconds, a dialogue box will appear (Figure 11.2). Click on **OK** to return to the **Generate Scenario** screen.

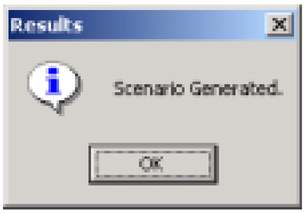
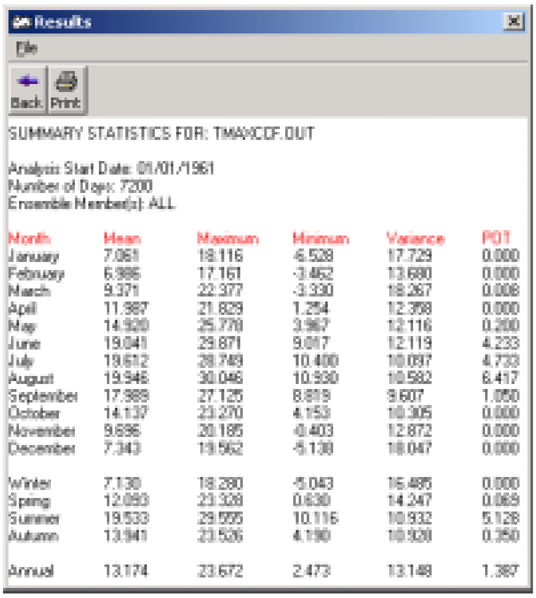


Figure 11.2 The Scenario Generated dialogue box.

11.3 Blogsville example (temperature)

For the Blogsville example, the **Generate Scenario** operation was applied twice. First predictors from the GCMX experiment for the period 1960–1989 were used to downscale current climate forcing. Figure 11.3 shows the **Results** screen for this time–slice, obtained from the **Analyse Model Output** operation (see Section 9).



SUMMARY STATISTICS FOR: TMAXCCF.OUT

Analysis Start Date: 01/01/1961
Number of Days: 7200
Ensemble Member(s): ALL

Month	Mean	Maximum	Minimum	Variance	POT
January	7.061	18.116	-6.528	17.729	0.000
February	6.986	17.161	-3.462	13.680	0.000
March	9.371	22.377	-3.330	18.267	0.008
April	11.987	21.829	1.254	12.358	0.000
May	14.920	25.778	3.967	12.116	0.200
June	19.041	29.871	9.017	12.119	4.233
July	19.612	28.749	10.400	10.097	4.733
August	19.946	30.046	10.930	10.582	6.417
September	17.989	27.125	8.819	9.607	1.050
October	14.137	23.270	4.153	10.305	0.000
November	9.696	20.185	-0.403	12.872	0.000
December	7.343	19.562	-5.138	18.047	0.000
Winter	7.130	18.280	-5.043	16.485	0.000
Spring	12.093	23.328	0.630	14.247	0.069
Summer	19.533	29.555	10.116	10.932	5.128
Autumn	13.941	23.526	4.190	10.928	0.350
Annual	13.174	23.672	2.473	13.148	1.387

Figure 11.3 Results for the Blogsville example using GCMX predictors (1960–1989).

Using the **Compare Results** operation, it is possible to compare the variance of maximum temperatures at Blogsville downscaled using observed (NCEP) and GCMX predictor variables. For example, Figure 11.4 shows that GCMX predictors generally over–estimate the variance of maximum temperatures with respect to the NCEP predictors. (Note that results for individual SDSM runs or ensemble members will differ slightly — even when using the same model parameters and predictors — due to the stochastic component of the downscaling.)

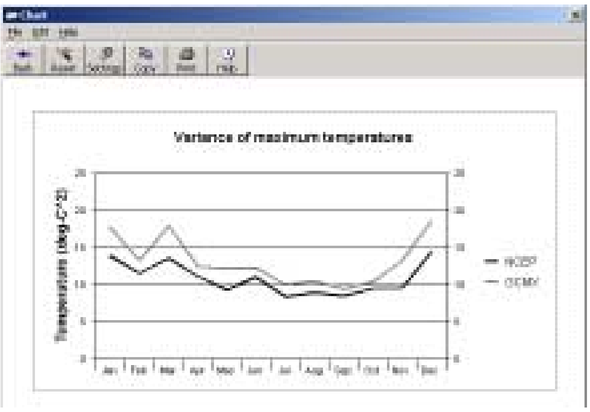


Figure 11.4 Monthly variance of maximum temperatures at Blogsville for the current climate downscaled using observed (NCEP) predictors (1961–1990) and GCMX predictors (1960–1989).

The second example of the **Generate Scenario** operation uses GCMX predictors for the period 2080–2099. As before, Figure 11.5 shows the **Results** screen for this scenario, obtained from the **Analyse Model Output** operation.

Results

File

Back Print

SUMMARY STATISTICS FOR: TMAXFCF.OUT

Analysis Start Date: 01/01/1961
Number of Days: 7200
Ensemble Member(s): ALL

Month	Mean	Maximum	Minimum	Variance	POT
January	7.171	19.104	-6.929	20.132	0.000
February	8.035	18.379	-3.660	14.847	0.000
March	9.384	21.471	-3.220	17.011	0.000
April	12.543	23.710	1.753	13.601	0.000
May	16.450	27.750	4.219	15.606	1.208
June	19.873	31.425	8.893	13.270	8.150
July	21.701	30.404	12.034	10.120	15.433
August	21.887	31.568	12.641	10.350	16.842
September	19.707	28.783	10.027	9.476	4.042
October	15.883	24.872	6.128	10.497	0.092
November	11.581	22.503	-0.553	14.729	0.008
December	8.263	20.099	-5.850	17.197	0.000
Winter	7.823	19.194	-5.483	17.392	0.000
Spring	12.792	24.310	0.917	15.406	0.406
Summer	21.154	31.132	11.123	11.247	13.475
Autumn	15.724	25.386	5.201	11.567	1.381
Annual	14.373	25.006	2.940	13.903	3.815

Figure 11.5 Maximum temperatures for Blogsville using GCMX predictors (2080–2099).

Once again, by using the **Compare Results** operation, it is possible to compare the frequency of “hot” days at Blogsville under current (1960–1989) and future (2080–2099) climate forcing. For example, Figure 11.6 shows a significant

increase in the frequency of hot–days in summer, most noticeably in the month of August. The downscaling also indicates that hot–days could begin to appear in May by the end of the 21st century.

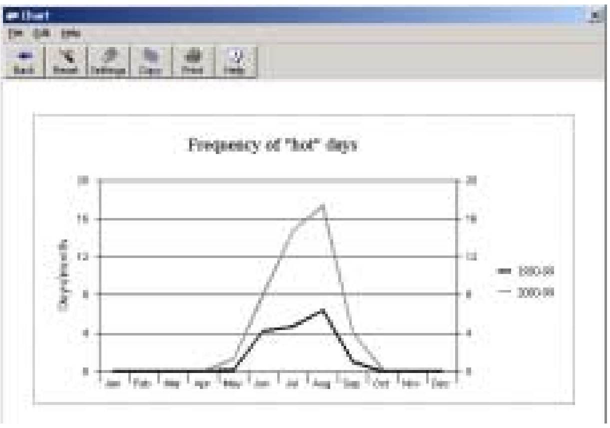
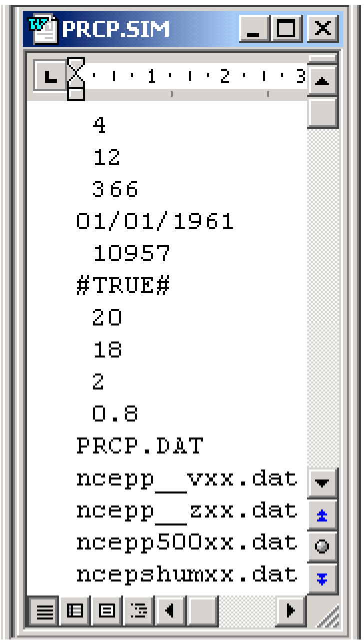


Figure 11.6 Monthly frequency of “hot” days (>25ºC) at Blogsville downscaled using GCMX predictors under current (1960–1989) and future (2080–2099) climate forcing.

11.4 Blogsville example (precipitation)

Precipitation downscaling is necessarily more problematic than temperature, because daily precipitation amounts at individual sites are relatively poorly resolved by regional–scale predictors, and because precipitation is a conditional process (i.e., both the occurrence and amount processes must be specified). Figure 11.7 shows a *.SIM file used to downscale daily precipitation using observed (NCEP) predictors.




```
[1] Number of predictor variables
[2] Number of regression models used
[3] Maximum number of days in a year
[4] Start date
[5] Number of days simulated
[6] Conditional or unconditional predictand
[7] Number of ensemble members
[8] Variance inflation parameter
[9] Transformation code
[10] Bias correction parameter
[11] Predictand file name
[12 onward] Predictor file name(s).
      “      “      “      “
      “      “      “      “
      “      “      “      “
      “      “      “      “
```

Figure 11.7 The *.SIM file for downscaling precipitation at Blogsville, 1961–1990.

Figure 11.7 shows that four predictors were employed (line 1), to simulate 12 months (line 2), using calendar years (line 3), beginning in 01/01/1961 (line 4) and lasting 10957 days (line 5). The model was conditional (#TRUE#, line 6), had 20 ensemble members (line 7), variance inflation (line 8), a fourth root transformation of the predictand (line 9) and bias correction (line 10). The four predictors were *p500*, *shum*, *p__v* and *p__z* (lines 12 onwards).

With the above specifications, the **Weather Generator** was used to downscale observed (NCEP) predictors, and **Generate Scenario** to downscale GCMX predictors representing the current climate. (Note that a **Year Length** of 366 days was checked in **Settings** when working with NCEP, and 360 when using GCMX predictors). Downscaled scenarios were evaluated, firstly using the **Analyse Model Output**, and then **Compare Results**. Figure 11.8 shows, for example, that the downscaling produced similar monthly mean 24–hour totals under observed (NCEP) and GCMX forcing for the current climate.

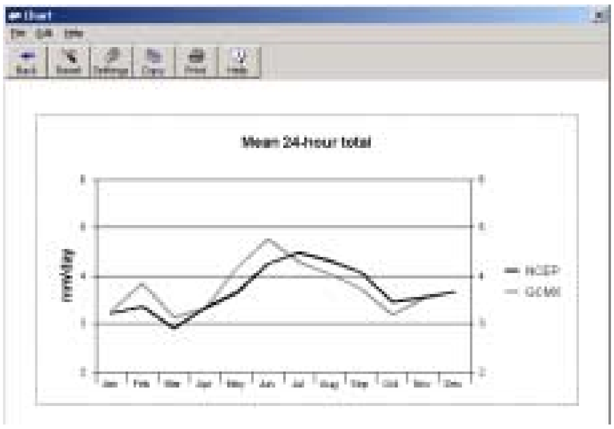


Figure 11.8 Monthly mean daily precipitation totals at Blogsville for the current climate downscaled using observed (NCEP) predictors (1961–1990) and GCMX predictors (1960–1989).

The **Generate Scenario** operation was implemented for a second time using GCMX predictors representing future (2080–2099) climate forcing. Figures 11.9 and Figure 11.10 compare selected outputs of this scenario with equivalent results from the current (1960–1989) climate downscaling using GCMX predictors.

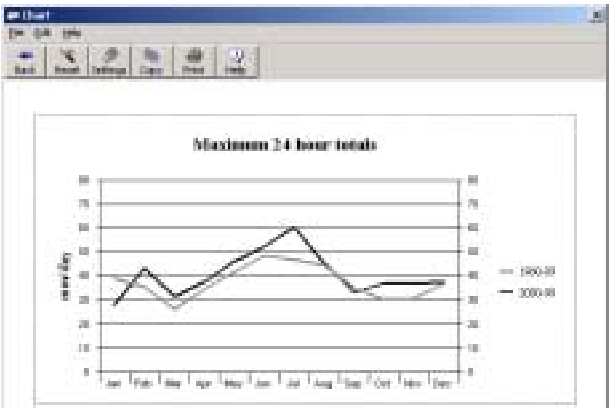


Figure 11.9 Monthly maximum daily precipitation totals at Blogsville downscaled using GCMX predictors under current (1960–1989) and future (2080–2099) forcing.

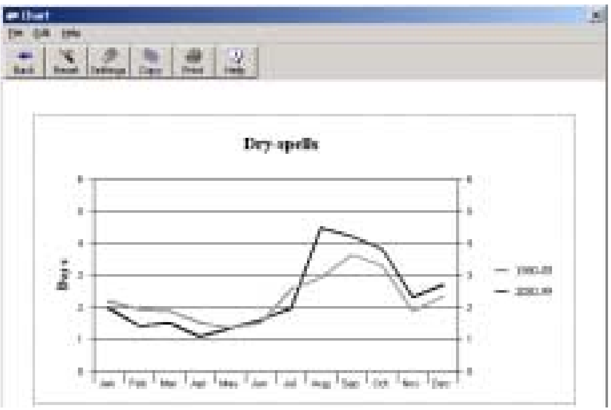


Figure 11.10 Monthly mean dry–spell lengths at Blogsville downscaled using GCMX predictors under current (1960–1989) and future (2080–2099) forcing.

The exemplar results presented in Figures 11.9 and 11.10 indicate a shift towards more intense summer storms, inter–spaced by longer dry–spells in late summer and autumn. Conversely, maximum winter intensities do not change significantly, but the average duration of dry–spells marginally decreases.

BIBLIOGRAPHY

This Bibliography provides references to papers containing full technical details of SDSM, along with example case studies. Additional background material is also recommended for other downscaling methods, as well as selected review papers in which various downscaling methods have been compared. Strongly recommended texts are denoted by asterisks (*****).

Technical basis and application of SDSM

Conway, D., Wilby, R.L. and Jones, P.D. 1996. Precipitation and air flow indices over the British Isles. *Climate Research*, **7**, 169–183.

Hassan, H., Aramaki, T., Hanaki, K., Matsuo, T. and Wilby, R.L. 1998. Lake stratification and temperature profiles simulated using downscaled GCM output. *Journal of Water Science and Technology*, **38**, 217–226.

*****Jones, P.D., Hulme, M. and Briffa, K.R. 1993. A comparison of Lamb circulation types with an objective classification scheme. *International Journal of Climatology*, **13**, 655–663.

Narula, S.C. and Wellington, J.F. 1977. An algorithm for linear regression with minimum sum of absolute errors. *Applied Statistics*, **26**, 106–111.

Rubinstein, R.Y. 1981. *Simulation and the Monte Carlo method*. Wiley, New York.

*****Wilby, R.L. and Dettinger, M.D. 2000. Streamflow changes in the Sierra Nevada, CA simulated using a statistically downscaled General Circulation Model scenario of climate change. In: McLaren, S.J. and Kniveton, D.R. (Eds.), *Linking Climate Change to Land Surface Change*, Kluwer Academic Publishers, Netherlands, pp. 99–121.

Wilby, R.L., Dawson, C.W. and Barrow, E.M. 2001. SDSM – a decision support tool for the assessment of regional climate change impacts. *Environmental and Modelling Software*, in press.

Wilby, R.L., Hassan, H. and Hanaki, K. 1998b. Statistical downscaling of hydrometeorological variables using general circulation model output. *Journal of Hydrology*, **205**, 1-19.

NCEP re–analysis and Hadley Centre experiments

*****Gordon, C., Cooper, C, Senior, C., Banks, H., Gregory, J., Johns, T., Mitchell, J. and Wood, R. 2000. The simulation of SST, sea–ice extents and ocean heat transport in a version of the Hadley Centre coupled model without flux adjustments. *Climate Dynamics*, **16**, 147–168.

Hadley Centre 1998. *Climate Change and its Impacts: Some Highlights from the Ongoing UK Research Programme*. UK Met. Office Publication 12pp.

*****Johns, T.C., Carnell, R.E., Crossley, J.F., Gregory, J.M., Mitchell, J.F.B., Senior, C.A., Tett, S.F.B. and Wood, R.A. 1997. The Second Hadley Centre coupled

ocean–atmosphere GCM: Model description, spinup and validation. *Climate Dynamics*, **13**, 103–134.

Kalnay, E., Kanamitsu, M., Kistler, R. *et al.* 1996. The NCEP/NCAR 40–year reanalysis project. *Bulletin of the American Meteorological Society,* **77**, 437–471.

Mitchell, J.F.B. and Johns, T.C. 1997. On modification of global warming by sulphate aerosols. *Journal of Climate,* **10**, 245–267.

Mitchell, J.F.B., Johns, T.C., Gregory, J.M. and Tett, S. 1995. Climate response to increasing levels of greenhouse gases and sulphate aerosols. *Nature,* **376**, 501–504.

Tett, S.F.B, Johns, T.C. and Mitchell, J.F.B. 1997. Global and regional variability in a coupled AOGCM. *Climate Dynamics,* **13**, 303–323.

Downscaling (general)

Department of the Environment. 1996. *Review of the potential effects of climate change in the United Kingdom*. HMSO, London, 247 pp.

Hulme, M. Jiang, T. and Wigley, T.M.L. 1995. *SCENGEN, a Climate Change Scenario Generator: User Manual*. Climatic Research Unit, University of East Anglia, Norwich, UK, 38 pp.

IPCC–TGCIA. 1999. *Guidelines on the Use of Scenario data for Climate Impact and Adaptation Assessment*. Version 1. Prepared by Carter, T.R., Hulme, M., and Lal, M, Intergovernmental Panel on Climate Change, Task Group on Scenarios for Climate Impact Assessment, 69pp.

Karl, T.R., Wang, W.C., Schlesinger, M.E., Knight, R.W. and Portman, D. 1990. A method of relating general circulation model simulated climate to the observed local climate. Part I: Seasonal statistics. *Journal of Climate,* **3**, 1053–1079.

Klein, W.H. and Glahn, H.R. 1974. Forecasting local weather by means of model output statistics. *Bulletin of the American Meteorological Society,* **55**, 1217-1227.

*****Wilby, R.L. and Wigley, T.M.L. 2000. Precipitation predictors for downscaling: observed and General Circulation Model relationships. *International Journal of Climatology,* **20**, 641-661.

*****Wilby, R.L. and Wigley, T.M.L. 1997. Downscaling general circulation model output: a review of methods and limitations. *Progress in Physical Geography,* **21**, 530–548.

Xu, C.–Y. 1999. From GCMs to river flow: a review of downscaling methods and hydrologic modelling approaches. *Progress in Physical Geography*, **23**, 229–249.

Statistical downscaling methods

Bardossy, A. and Plate, E.J. 1992. Space–time model for daily rainfall using atmospheric circulation patterns. *Water Resources Research,* **28**, 1247–1259.

Burger, G. 1996. Expanded downscaling for generating local weather scenarios. *Climate Research*, **7**, 111–128.

Crane, R.G. and Hewitson, B.C. 1998. Doubled CO2 precipitation changes for the Susquehanna Basin: downscaling from the GENESIS general circulation model. *International Journal of Climatology*, **18**, 65–76.

Hay, L.E., McCabe, G.J., Wolock, D.M. and Ayers, M.A. 1991. Simulation of precipitation by weather type analysis. *Water Resources Research*, **27**, 493–501.

Huth, R. 1999. Statistical downscaling in central Europe: evaluation of methods and potential predictors. *Climate Research*, **13**, 91–101.

***Kilsby, C.G., Cowpertwait, P.S.P., O'Connell, P.E. and Jones, P.D. 1998. Predicting rainfall statistics in England and Wales using atmospheric circulation variables. *International Journal of Climatology*, **18**, 523-539.

McCabe, G.J. and Dettinger, M.D. 1995. Relations between winter precipitation and atmospheric circulation simulated by the Geophysical Fluid Dynamic Laboratory General Circulation Model. *International Journal of Climatology*, **15**, 625-638.

Schubert, S. and Henderson-Sellers, A. 1997. A statistical model to downscale local daily temperature extremes from synoptic-scale atmospheric circulation patterns in the Australian region. *Climate Dynamics,* **13**, 223-234.

Wilby, R.L. 1997. Non–stationarity in daily precipitation series: implications for GCM downscaling using atmospheric circulation indices. *International Journal of Climatology*, **17**, 439–454.

Weather generators

Nicks, A.D., Lane, L.J. and Gander, G.A. 1995. Weather generator. In: Flanagan, D.C. and Nearing, M.A. (Eds.), *USDA–Water Erosion Prediction Project: Hillslope Profile and Watershed Model Documentation*. USDA–ARS National Soil Erosion Research Laboratory Report No. 10, West Lafayette, IN, USA.

Richardson, C.W. 1981. Stochastic simulation of daily precipitation, temperature and solar radiation. *Water Resources Research*, **17**, 182–190.

Semenov, M.A. and Barrow, E.M. 1997. Use of a stochastic weather generator in the development of climate change scenarios. *Climatic Change*, **35**, 397–414.

Wilks, D.S. 1992. Adapting stochastic weather generation algorithms for climate change studies. *Climatic Change*, **22**, 67–84.

Wilks, D.S. 1999. Multisite downscaling of daily precipitation with a stochastic weather generator. *Climate Research*, **11**, 125–136.

***Wilks, D.S. and Wilby, R.L. 1999. The weather generation game: a review of stochastic weather models. *Progress in Physical Geography*, **23**, 329–357.

Dynamical downscaling methods

Christensen, J.H., Machenhauer, B., Jones, R.G., Schär, C., Ruti, P.M., Castro, M. and Visconti, G. 1997. Validation of present-day regional climate simulations over

Europe: LAM simulations with observed boundary conditions. *Climate Dynamics*, **13**, 489-506.

Giorgi, F. and Mearns, L.O. 1991. Approaches to the simulation of regional climate change. A review. *Rev. Geophys.*, **29**, 191-216.

***Giorgi, F. and Mearns, L.O. 1999. Introduction to special section: Regional climate modeling revisited. *Journal of Geophysical Research*, **104**, 6335–6352.

McGregor, J.J. 1997. Regional climate modelling. *Meteorological and Atmospheric Physics*, **63**, 105–117.

Downscaling comparisons

Charles, S.P., Bates, B.C., Whetton, P.H. and Hughes, J.P. 1999. Validation of downscaling models for changed climate conditions: case study of southwestern Australia. *Climate Research*, **12**, 1–14.

Hay, L.E., Wilby, R.L. and Leavesley, G.H. 2000. A comparison of delta change and downscaled GCM scenarios for three mountainous basins in the United States. *Journal of the American Water Resources Association*, **36**, 387–397.

***Kidson, J. W. and Thompson, C.S. 1998. A comparison of statistical and model–based downscaling techniques for estimating local climate variations. *Journal of Climate*, **11**, 735–753.

Mearns, L.O., Bogardi, I., Giorgi, F., Matyasovszky, I. and Palecki, M. 1999a. Comparison of climate change scenarios generated from regional climate model experiments and statistical downscaling. *Journal of Geophysical Research*, **104**, 6603–6621.

Mearns, L.O., Mavromatis, T., Tsvetsinskaya, E., Hays, C. and Easterling, W. 1999b. Comparative responses of EPIC and CERES crop models to high and low spatial resolution climate change scenarios. *Journal of Geophysical Research*, **104**, 6623–6646.

***Murphy, J. 1999. An evaluation of statistical and dynamical techniques for downscaling local climate. *Journal of Climate*, **12**, 2256–2284.

Murphy, J. 2000. Predictions of climate change over Europe using statistical and dynamical downscaling methods. *International Journal of Climatology*, **20**, 489–501.

Takle, E.S., Gutowski, Jr., W.J., Arritt, R.W., Pan, Z., Anderson, C.J., Silva, R., Caya, D., Chen, S.-C., Christensen, J.H., Hong, S.-Y., Juang, H.-M. H., Katzfey, J.J., Lapenta, W.M., Laprise, R., Lopez, P., McGregor, J. and Roads, J.O. 1999. Project to Intercompare Regional Climate Simulations (PIRCS): Description and initial results. *Journal of Geophysical Research*, **104**, 19,443-19,462.

Wilby, R.L., Hay, L.E. and Leavesley, G.H. 1999. A comparison of downscaled and raw GCM output: implications for climate change scenarios in the San Juan River Basin, Colorado. *Journal of Hydrology*, **225**, 67–91.

Wilby, R.L., Hay, L.E., Gutowski, W.J., Arritt, R.W., Takle, E.S., Pan, Z., Leavesley, G.H. and Clark, M.P. 2000. Hydrological responses to dynamically and statistically downscaled climate model output. *Geophysical Research Letters*, **27**, 1199–1202.

*****Wilby, R.L., Wigley, T.M.L., Conway, D., Jones, P.D., Hewitson, B.C., Main, J. and Wilks, D.S. 1998a. Statistical downscaling of General Circulation Model output: a comparison of methods. *Water Resources Research*, **34**, 2995–3008.

Winkler, J.A., Palutikof, J.P., Andresen, J.A. and Goodess, C.M. 1997. The simulation of daily temperature series from GCM output. Part II: Sensitivity analysis of an empirical transfer function methodology. *Journal of Climate*, **10**, 2514-2532.

APPENDIX 1: EHNHANCEMENTS SINCE VERSION 2.1

Development of the prototype SDSM (Version 2.1) in November 2000, was sponsored by A Consortium for the Application of Climate Impact Assessments (ACACIA), and the Canadian Institute for Climate Studies (CICS) through the Climate Change Action Fund. This version is described by Wilby et al. (2001).

Development of SDSM Version 2.2 was sponsored entirely by the UK Environment Agency through the National Centre for Risk Analysis and Options Appraisal. Since SDSM Version 2.1 the following major enhancements have been made:

A.1.1 User interface

The User interface has been significantly upgraded to improve the versatility of the software and ease of use. Primary operations of SDSM have been arranged sequentially through drop–down menus and **Back** and **Next** buttons at the head of each screen. Where possible, generic tasks and options have been placed in global **Settings** and **Advanced Settings** screens that can be accessed throughout the software. Options available in individual screens have been rationalised where possible (e.g., **Calibrate Model** has been simplified). The general presentation of the software has been significantly improved throughout.

A.1.2 Quality control and data transformation

It is recognised that meteorological data are seldom 100% complete and/or entirely free of errors. Simple **Quality Control** checks are now available within SDSM. These checks enable the identification of gross data errors, missing data codes and outliers prior to model calibration.

In many instances it is helpful to transform predictors and/or predictands prior to model calibration. The inclusion of a **Transform** facility allows specified data files to be subjected to selected transformations (e.g., logarithm, power, inverse, lag, binomial conversion, etc). The lag facility supersedes the autocorrelation functions that were previously hard–coded in the former **Calibrate Model** screen.

A.1.3 Screen variables operation

The **Screen Variables** operation has undergone considerable change to assist Users in the selection of appropriate predictor variables. This screen now provides a full correlation matrix and partial correlations. A **Scatterplot** facility allows visual examination of individual predictor/predictand combinations for specified months, seasons or years.

A.1.4 Statistical analyses of model output

Additional diagnostic statistics have been included to provide the User with more information about the properties of observed data and model output. The required tests are listed in a **Statistics Selection** menu. Since Version 2.1, the following tests have been added to the list: median, peak(s) over/below threshold(s),

percentiles, inter–quartile range, lag–1 serial correlation, skewness, mean wet–/dry–spell length, and maximum wet–/dry–spell length. The results of the specified tests are written to a User–defined file for subsequent graphing (see below).

A.1.5 Visualisation of data

The previous version of SDSM had no facility for charting observed data or model output. SDSM Version 2.2 includes scatterplots (within the **Screen Variables** operation), bar charts and line charts (within the **Compare Results** operation). The User can now customise the style of the graphing by editing chart fonts, colour schemes, legend titles, scales, tick marks and placement. The **Compare Results** facility also enables side–by–side bar or line chart comparisons of monthly statistics derived from two data sets (e.g., model versus observed, current versus future climate scenario, etc.).

A.1.6 On–line help

The on–line **Help** of SDSM has been significantly upgraded. Placing the mouse–pointer over a button or screen options automatically generates a Help "balloon". Clicking on the Help button at the top of each screen generates context sensitive information. Selecting **Help Contents** from the Help drag–down menu leads the User to an itemised list of SDSM operations. From here it is now possible to search the **Help Index** by key term(s). The explanation provided under each entry has been expanded and clarified.

A.1.7 Miscellaneous changes

In addition to the above major enhancements, the capability of SDSM has been improved in other ways. For example, the model building process has been simplified by forcing a clearer divide between **Unconditional** and **Conditional** processes. The variance inflation and transformation applied (if any) to downscaled processes can now be specified by the User under **Advanced Settings**. User preferences (e.g., date ranges, model process, variance inflation, level of bias correction, etc.) can all be retained for later use in a **Settings File** that automatically loads on software start–up.

APPENDIX 2: FREQUENTLY ASKED QUESTIONS

The following generic and specific questions are arranged in the order in which they would be typically encountered during a downscaling procedure.

Q. Do I need to perform any re–gridding or normalisation of the predictor variables?

No. These tasks have already been performed for the UKSDSM data set released with the software. All UK data have been re–gridded to a standard co–ordinate system (2.5º latitude × 3.75º longitude), and normalised with respect to the 1961–1990 climatology. The User must simply select the nearest grid box(es) to the site in question.

Q. Can I use observational data that lie outside the standard period 1961 to 2000?

No. Observed predictor variables for the UKSDSM data set were obtained from NCEP and normalised only for the period 1961 to 2000. Station meteorological data prior to 1st January 1961 or after 31st December 2000 will have no associated predictor variables. The software also assumes that meteorological data provided by the User commences on 1st January 1961 (i.e., has the same start date as the predictors). If this is not the case, the User should pad the station data with the **Missing Data Identifier**.

Q. How important is the selection of predictor variables?

Identifying sensible predictor–predictand relationships is the most critical procedure in all statistical downscaling methods. The **Screen Variables** screen is designed to assist the User in the choice of appropriate downscaling predictor variables for model calibration via seasonal correlation analysis, partial correlation analysis, and scatterplots. Ultimately, however, the User must decide whether or not the identified relationships are physically sensible for the site(s) in question.

Q. How can I determine if I have chosen the correct predictor variables for the predictands that I require?

The correlation statistics and *p* values indicate the strength of the *association* between two variables. Higher correlation values imply a higher degree of association. Smaller *p* values indicates that this association is less likely to have occurred by chance. A p value <0.05 is routinely used as the cut-off, so a *p* value of 0.37 would indicate that the predictor–predictand correlation is likely to be due to chance. However, even if $p < 0.05$ the result can be statistically significant but *not* be of practical significance — there's a difference!

Even if a high correlation and low *p* value is returned, the **Scatterplot** indicates whether this result is due to a few outliers, or is a potentially useful downscaling relationship. The **Scatterplot** may also reveal that one (or both) of the

variables should by modified using the **Transform** operation, to linearise the relationship.

Q. How do I build "memory" (autocorrelation) into the model?

There are TWO ways of incorporating memory in the downscaling model. Firstly, it is possible to use the **Transform** facility to create lagged predictands. For example, TMAXlag1 (i.e., the value of TMAX on preceding days) is used as a predictor of TMAX when calibrating the model. Secondly, the **Transform** facility is to create a selection of lagged predictor variables. For example, if V is correlated with PRCP, then consider Vlag-1, Vlag-2, Vlag-3, etc. It is better practise to employ lagged predictors, not lagged predictands in the model calibration. This makes greater physical sense and avoids many problems when it comes to generating synthetic data.

Q. What does the Event Threshold parameter (in Settings) do?

The **Event Threshold** parameter specifies the boundary between the two states in a **Conditional** process model. For example, if the Conditional process is precipitation, changing the Event Threshold from 0 to 0.3 will result in more "dry" days and fewer "wet" days (a simple way of classifying "trace" rainfall days as dry days). Therefore, different values for the Event Theshold will yield different results in **Screen Variables** (correlation values and scatterplots are both affected), will produce different parameters in **Calibrate Model**, and different results from the two **Analyse** operations. Note, however, that the **Weather Generator** and **Generate Scenario** operations will produce values in the range 0 to 0.3 even if the threshold is set at 0.3

Q. What are the advantages and disadvantages of using the monthly, seasonal or annual Model Type in Calibrate Model?

The **Model Type** button in **Calibrate Model** determines whether individual downscaling models will be calibrated for each calendar month, climatological season or entire year. The monthly button should be selected whenever the predictand is known to have a strong seasonal cycle, noting that even the annual button can produce the same result *provided* that one or more predictor variables have strong seaonality. Annual models are more parsimonious in the sense that they have only one set of regression weights instead of twelve in the case of the monthly models. Seasonal models might be used in situations where data are too sparse at the monthly level for model calibration, for example, a low incidence of summer precipitation rainfall.

Q. I am trying to model precipitation and have chosen the fourth root transformation in Advanced Settings. What else must I do?

Nothing! The fourth root button in **Advanced Settings** tells the software that this transformation is to be used throughout (including calibration, weather generation and scenario generation). If checked, there's no need to apply any further transformations as this is all backed out automatically. So when calibrating the model with fourth root checked, you should supply the model with untransformed rainfall data, making sure that the **Conditional** process button is checked.

Q. Is it OK to model precipitation as an unconditional process?

As a general rule, precipitation should be modelled as a **Conditional** process. It does not make much sense to neglect the occurrence process (i.e., sequences of wet or dry days are first modelled, then the amounts if it is a wet day). If you are being swayed by higher R-sq values of an **Unconditional** model during calibration, beware, the result is probably seriously biased by the large number of zero values entered in the multiple regression. Remember, daily precipitation amount is the most problematic daily variable to downscale.

Q. When I use the Weather Generator I get unrealistically large maximum daily (precipitation) values. What's going wrong?

Unrealistically large values generally imply that the variance inflation and/or bias correction in **Advanced Settings** are too high.

Q. Why do I get slightly different results every time I run the Weather Generator (with the same inputs)?

Even with the same inputs (i.e., *.PAR file, **Settings** and data period) the **Weather Generator** (and **Generate Scenario**) operation is not expected to produce identical results. This is because of the stochastic (random) component that is applied to each downscaled series to compensate for the fact that the deterministic component of the model (due to the chosen predictor variables) does not explain all of the observed variance. Differences between individual runs and/or **Ensemble Members** is likely to be greater for poorly determined predictands such as precipitation than in better constrained predictands such as temperature.

Q. Does SDSM produce realistic results for multiple sites? Also, what if I'm interested in preserving relationships between variables?

Both of these questions are the subject of ongoing research. However, results from previous studies suggest that regression–based downscaling does preserve a significant part of observed inter–site correlations *provided* that models calibrated on a site by site basis are forced by a common set of predictors. In other words, inter–site correlations are implicitly reproduced by virtue of correlated predictor variables, rather than by the model structure.

Preliminary tests of inter–variable correlations produced by SDSM (e.g., between downscaled precipitation and temperature series) indicate that inter–annual variations in the strength of relationships are preserved, but there can be differences between the model and observations in individual months. Once again, it is suspected that inter–variable relationships are implicitly preserved by virtue of commonality in the predictor variables used to downscale each predictand.

However, if required, it is relatively straightforward to explicitly condition one predictand on another (e.g., daily precipitation occurrence might be used to condition maximum temperatures). In this case, the conditioning variable (precipitation occurrence) would be entered as a predictor during model calibration.

Q. I’ve calibrated my model. How do I now produce values of PRCP, TMAX or TMIN using GCM data?

Provided you have produced a **.PAR* file via **Calibrate Model**, the software will automatically know what predictors are needed. Of course you may need to transform some of the GCM files if this was done for calibration. For example, if *p__z* was transformed to *psqz* and then used to train the model, the same transformation should be applied to the equivalent GCM file(s) used for downscaling climate model output. In any event, be sure to employ the SDSM nomenclature for all transformed files.

Q. Why do I keep getting an error message when I use GCM data?

The most likely explanation is that the **Year Length** in **Settings** has not been set correctly with respect to the number of days in the GCM simulation. For example, HadCM2 and HadCM3 have year lengths of 360 days, whereas CGCM1 has 365 days in every year (i.e., no leap years).

Q. What’s the best way of handling SDSM files outside the software?

All SDSM output files are written in ASCII format and, therefore, accessible by any word processor. Model results (*.OUT files) are tab–delimited if the number of **Ensemble Members** is greater than one, and, as such, can be imported into Excel for further analysis or graphing.

Q. I’ve looked at the predictor variable files and the values only range between +/-5. Is there something wrong with the data?

No. Most predictor variables (NCEP and GCM) are normalised using their respective 1961–1990 means and standard deviations. The result is that each predictor variable is dimensionless, and will typically vary between –5 and +5.

GLOSSARY

Where appropriate, the following definitions were drawn from the Glossary of terms in the Summary for Policymakers, A Report of Working Group I of the Intergovernmental Panel on Climate Change, and the Technical Summary of the Working Group I Report.

Terms in *italics* are found elsewhere in this Glossary.

Aerosols Airborne solid or liquid particles, with a typical size between 0.01 and 10μm that reside in the atmosphere for at least several hours. Aerosols influence the *climate* directly through scattering and absorbing radiation, and indirectly through the formation and optical properties of clouds.

Airflow (index) Trigonometric measures of atmospheric circulation obtained from surface pressure or geopotential height fields. Commonly derived indices include *vorticity*, *zonal flow*, *meridional flow*, and *divergence*. Certain indices have been used to replicate subjective classifications of daily *weather patterns*, or as predictor variables in statistical *downscaling* schemes.

Anthropogenic Resulting from, or produced by, human beings.

Atmosphere The gaseous envelope surrounding the Earth, comprising almost entirely of nitrogen (78.1%) and oxygen (20.9%), together with several trace gases, such as argon (0.93%) and *greenhouse gases* such as carbon dioxide (???%).

Black box Describes a system or model for which the inputs and outputs are known, but intermediate processes are either unknown or unprescribed. See *regression*.

Climate The "average weather" described in terms of the mean and variability of relevant quantities over a period of time ranging from months to thousands or millions of years. The classical period is 30 years, as defined by the World Meteorological Organisation (WMO).

Climate change Statistically significant variation in either the mean state of the *climate*, or in its variability, persisting for an extended period (typically decades or longer). Climate change may be due to natural internal processes or to *external forcings*, or to persistent *anthropogenic* changes in the composition of the atmosphere or in land use.

Climate model A numerical representation of the climate system based on the physical, chemical and biological properties of its components, their interactions and feedback processes, and accounting for all or some its known properties.

Climate prediction An attempt to produce a most likely description or estimate of the actual evolution of the climate in the future, e.g. at seasonal, inter–annual or long–term time scales.

Climate projection A projection of the response of the climate system to emission or concentration scenarios of *greenhouse gases* and *aerosols*, or *radiative forcing* scenarios, often based on simulations by *climate models*. As such climate projections are based on assumptions concerning future socio–economic and technological developments.

Climate scenario A plausible and often simplified representation of the future climate, based on an internally consistent set of climatological relationships, that has

been constructed for explicit use in investigating the potential consequences of anthropogenic *climate change*.

Climate variability Variations in the mean state and other statistics (such as standard deviations, the occurrence of extremes, etc.) of the climate on all temporal and spatial scales beyond that of individual weather events.

Conditional process A mechanism in which an intermediate state variable governs the relationship between regional forcing and local weather. For example, local precipitation amounts are conditional on wet–day occurrence (the state variable), which in turn depends on regional–scale predictors such as atmospheric humidity and pressure.

Deterministic A process, physical law or model that returns the same predictable outcome from repeat experiments when presented with the same initial and boundary conditions, in contrast to *stochastic* processes.

Domain A fixed region of the Earth's surface and over-lying atmosphere represented by a *Regional Climate Model*. Also, denotes the grid box(es) used for statistical *downscaling*. In both cases, the downscaling is accomplished using pressure, wind, temperature or vapour information supplied by a host GCM.

Divergence If a constant volume of fluid has its horizontal dimensions increased it experiences divergence and, by conservation of mass, its vertical dimension must decrease.

Downscaling The development of climate data for a point or small area from regional climate information. The regional climate data may originate either from a *climate model* or from observations. Downscaling models may relate processes operating across different time and/or space scales.

Dynamical See *Regional Climate Model*.

Emission scenario A plausible representation of the future development of emissions of substances that are potentially radiatively active (e.g. *greenhouse gases*, *aerosols*), based on a coherent and internally consistent set of assumptions about driving forces and their key relationships.

Ensemble (member) A set of simulations (members) in which a deterministic *climate model* is run for multiple *climate projections*, each with minor differences in the initial or boundary conditions. Conversely, *weather generator* ensemble members differ by virtue of random outcomes of successive model simulations. In either case, ensemble solutions can be grouped and then compared with the ensemble mean to provide a guide to the *uncertainty* associated with specific aspects of the simulation.

External forcing A set of factors that influence the evolution of the climate system in time (and excluding natural internal dynamics of the system). Examples of external forcing include volcanic eruptions, solar variations and human–induced forcings such as changing the composition of the atmosphere and land use change.

Extreme weather event An event that is rare within its statistical reference distribution at a particular place. Definitions of "rare" vary from place to place (and from time to time), but an extreme event would normally be as rare or rarer than the 10th or 90th percentile.

General Circulation Model (GCM) A three–dimensional representation of the Earth's atmosphere using four primary equations describing the flow of energy (first

law of thermodynamics) and momentum (Newton's second law of motion), along with the conservation of mass (continuity equation) and water vapour (ideal gas law). Each equation is solved at discrete points on the Earth's surface at fixed time intervals (typically 10–30 minutes), for several layers in the atmosphere defined by a regular *grid* (of about 200km resolution). Couple ocean–atmosphere general circulation models (O/AGCMs) also include ocean, land–surface and sea–ice components. See *climate model*.

Geopotential height The work done when raising a body of unit mass against gravity (i.e., acceleration due to gravity at a given level in the atmosphere multiplied by distance) divided by the value of gravity at the Earth's surface.

Greenhouse gas Gaseous constituents of the atmosphere, both natural and anthropogenic, that absorb and emit radiation at specific wavelengths within the spectrum of infrared radiation emitted by the Earth's surface, the atmosphere and clouds. The primary greenhouse gases are water vapour (H_2O), carbon dioxide (CO_2), nitrous oxide (N_2O), methane (CH_4), and ozone (O_3).

Grid The co–ordinate system employed by *GCM* or *RCM* to compute three–dimensional fields of atmospheric mass, energy flux, momentum and water vapour. The grid spacing determines the smallest features that can be realistically resolved by the model. Typical resolutions for GCMs are 200km, and for RCMs 20–50km.

Meridional flow An atmospheric circulation in which the dominant flow of air is from north to south, or from south to north, across the parallels of latitude, in contrast to *zonal flow*.

NCEP The acronym for the National Center for Environmental Prediction. The source of re–analysis (climate model assimilated) data widely used for dynamical and statistical *downscaling* of the current climate.

Normalisation A statistical procedure involving the standardisation of a data set (by subtraction of the mean and division by the standard deviation) with respect to a predefined control period. The technique is widely used in statistical *downscaling* to reduce systematic biases in the mean and variance of climate model output.

Parameter A numerical value representing a process or attribute in a model. Some parameters are readily measurable climate properties; others are known to vary but are not specifically related to measurable features. Parameters are also used in climate models to represent processes that poorly understood or resolved.

Predictand A variable that may be inferred through knowledge of the behaviour of one or more *predictor* variables.

Predictor A variable that is assumed to have predictive skill for another variable of interest, the *predictand*. For example, day–to–day variations in atmospheric pressure may be a useful predictor of daily rainfall occurrence.

Radiative forcing The change in net vertical irradiance (expressed as Watts per square metre) at the *tropopause* due to an internal change or a change in the *external forcing* of the climate system, such as, for example, a change in the concentration of carbon dioxide, or the output of the Sun.

Random See *stochastic*.

Re–gridding A statistical technique used to project one co–ordinate system onto another, and typically involving the interpolation of climate variables. A necessary

pre–requisite to most statistical *downscaling*, because observed and climate model data are seldom archived using the same *grid* system.

Regional Climate Model (RCM) A three–dimensional, mathematical model that simulates regional scale climate features (of 20–50 km resolution) given time–varying, atmospheric properties modelled by a General Circulation Model. The RCM *domain* is typically "nested" within the three–dimensional *grid* used by a GCM to simulate large–scale fields (e.g. surface pressure, wind, temperature and vapour).

Regression A statistical technique for constructing empirical relationships between a dependent (*predictand*) and set of independent (*predictor*) variables. See also *black box, transfer function*.

Relative humidity A relative measure of the amount of moisture in the air to the amount needed to saturate the air at the same temperature expressed as a percentage.

Resolution The *grid* separation of a climate model determining the smallest physical feature that can be realistically simulated.

Scenario A plausible and often simplified description of how the future may develop based on a coherent and internally consistent set of assumptions about driving forces and key relationships. Scenarios may be derived from projections, but are often based on additional information from other sources, sometimes combined with a "narrative story–line".

Specific humidity The ratio of the mass of water vapour (in grams) to the mass of moist air (in kilograms) in a given volume of air.

Station The individual site at which meteorological measurements are systematically observed and recorded.

Stochastic A process or model that returns different outcomes from repeat experiments even when presented with the same initial and boundary conditions, in contrast to *deterministic* processes. See *weather generator*.

Transfer function A mathematical equation that relates a *predictor*, or set of predictor variables, to a target variable, the *predictand*. The predictor(s) and predictand represent processes operating at different temporal and/or spatial scales. In this case, the transfer function provides a means of *downscaling* information from coarse to finer resolutions.

Tropopause The boundary between the lowest part of the atmosphere, known as the troposphere, and the highly stratified region of the atmosphere, known as the stratosphere. The tropopause is typically located 10km above the Earth's surface.

Uncertainty An expression of the degree to which a value (e.g. the future state of the climate system) is unknown. Uncertainty can result from a lack of information or from disagreement about what is known or knowable. It can also arise from poorly resolved climate model parameters or boundary conditions.

Unconditional process A mechanism involving direct physical or statistical link(s) between a set of *predictors* and the *predictand*. For example, local wind speeds may be a function of regional airflow strength and *vorticity*.

Vorticity Twice the angular velocity of a fluid particle about a local axis through the particle. In other words, a measure of rotation of an air mass.

Weather generator A model whose stochastic (random) behaviour statistically resembles daily weather data at a location. Unlike *deterministic* weather forecasting models, weather generators are not expected to duplicate a particular weather sequence at a given time in either the past or the future. Most weather generators assume a link between the precipitation process and secondary weather variables such as temperature, solar radiation and humidity.

Weather pattern An objectively or subjectively classified distribution of surface (and/or upper atmosphere) meteorological variables, typically daily mean sea level pressure. Each atmospheric circulation pattern should have distinctive meteorological properties (e.g. chance of rainfall, sunshine hours, wind direction, air quality, etc). Examples of subjective circulation typing schemes include the European Grosswetterlagen, and the British Isles Lamb Weather Types.

Zonal flow An atmospheric circulation in which the dominant flow of air follows the lines of latitude (e.g. the westerlies), in contrast to *meridional flow*.